Sampo Sammalisto

Search for Genetic Variants Influencing Human Height

Publications of the National Public Health Institute A 1/2008

Department of Molecular Medicine,
National Public Health Institute
and
Department of Medical Genetics,
Faculty of Medicine, University of Helsinki

Helsinki, Finland 2008

Sampo Sammalisto

SEARCH FOR GENETIC VARIANTS INFLUENCING HUMAN HEIGHT

A C A D E M I C D I S S E R T A T I O N

To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in the Small Hall, University Main Building, on February 1st, at 12 noon.

National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2008

Helsinki University Biomedical Dissertations No 99

ISSN 1457-8433

Publications of the National Public Health Institute KTL A1 / 2008

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-759-5
ISSN 0359-3584
ISBN 951-740-760-1 (pdf)
ISSN 1458-6290 (pdf)

Kannen kuva - cover graphic: A living histogram from Conneticut State Agricultural College. Reprinted from Blakeslee, A.F., 1914, Corn and men. J. Hered. 5: 511-518 with permission.

Yliopintopaino
Helsinki 2008

Supervised by

Academy Professor Leena Peltonen-Palotie
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Adjunct Professor Markus Perola
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Reviewed by

Adjunct Professor Päivi Onkamo
University of Helsinki,
Division of Genetics,
Department of Biological and Environmental Sciences,
Helsinki, Finland

Adjunct Professor Pekka Uimari
Oy Jurilab Ltd.
Kuopio, Finland

Opponent

Harald Göring, Ph.D.
Southwest Fondation for Biomedical Research,
Department of Genetics,
San Antonio, Texas, USA

“Tosiasioiden tunnustaminen on kaiken viisauden alku”

J. K. Paasikivi (1870-1956)

To my family.

Sampo Sammalisto, Search for genetic variants influencing human height
Publications of the National Public Health Institute, A1/2008, 124 Pages
ISBN 951-740-759-5; 951-740-760-1 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

ABSTRACT

Human height is determined by a combination of genetic and environmental effects and in modern Western societies > 80% of the observed variation in height is determined by genetic factors. Height is a fundamental human trait that is associated with many socioeconomic and psychosocial factors and health measures, however little is known of the identity of the specific genes that influence height variation in the general population.

This study aims to identify the genetic variants that influence height in the general population by genome-wide linkage analysis in large collections of family samples from Australia, Europe and the United States. For this purpose, we carried out three genome-wide scans for adult height in 1) 1,417 individuals from 277 Finnish families, 2) 8,450 individuals from 3,817 families from Australia and Europe and 3) 9,306 individuals from 3,302 families from the United States.

In the Finnish families we found significant evidence for male-specific linkage to 1p21 (LOD = 4.25) as well as several suggestive linkages (LOD ≥ 2.0) implicating putative quantitative trait loci (QTL) for height on 4q35, 9p24, 13q12, 18q21 and 22q13. We followed up the 1p21 locus linkage finding with additional genotyping of single nucleotide polymorphism (SNP) markers in positional candidate genes and found an association to a functional variant on collagen 11 alpha 1 (COL11A1) in family-based association analyses (p=0.003) which we also replicated (p=0.03) in an independent Finnish population cohort (n=6,542) representative of the general population. From this population sample, we estimated that homozygosity for the minor allele of this SNP increased height by 1.1 cm in males and 0.6 cm in females and in total this SNP accounted for 0.1% of the population variance in height.

The Australian and European (Finnish, Danish, Dutch, Swedish and UK) families were ascertained for dizygotic twin (DZ) pairs and they were derived from the GenomEUtwin (www.genomeutwin.org) consortium twin cohorts. DZ twins provide special advantages for genetic studies of height because they share most of the environmental influences in critical periods of growth thus reducing within-pair environmental variance. In addition to the DZ twin pairs (n=6,602) for most families we also had data on additional family members increasing the total sample to 8,450

individuals. In these families we found significant evidence for linkage to stature on 8q21 (LOD=3.28) as well as several suggestive linkages (LOD ≥ 2.0) on 7p22, 20p13, 21q21 and Xq25. Also, at many of the loci restricting the analyses to DZ twins increased the evidence for linkage thus demonstrating the benefit of utilizing twin pairs for reducing environmental variance in genetic analyses. We also performed an independent genome-wide association (GWA) study with 317,000 SNP markers in 1,552 monozygotic twins where the most significant finding was also on 8q21-q24 further supporting the presence of height QTL at this locus.

The families from the United States (n=3,032) consisted of both African-American (n=1,628) and European-American families (n=1,404), therefore in addition to analyzing all families jointly we also stratified this sample according to self-reported ethnicity in order to reduce heterogeneity due to genetic and environmental sources. In these samples the strongest evidence for linkage was observed on 15q25 (LOD=3.0) and several other regions showed suggestive evidence for linkage (LOD ≥ 2.0) to stature on 11q23, 12q12, 15q26, 18q23 and 19q13. Of the suggestive linkage regions 12q12 is of particular interest since it overlaps with the high-mobility group at-hook 2 (HMGA2) gene which was recently reported for association with adult and childhood height in a large GWA study involving multiple populations. Also the linkage evidence on 18q23 overlaps well with the linkage observed in the Finnish families in this study and implicating this genomic region as potentially harbouring QTL that influence height across diverse populations.

In this study we identified one gene variant that influences human height, although this variant alone explains only 0.1% of height variation in the Finnish populations. We also showed converging evidence for the involvement of chromosome 8q21-q24 in the determination of adult height from two independent studies with complimentary genome-wide designs: family-based linkage and association using unrelated individuals. This study demonstrates that special stratification strategies such as performing sex-limited analyses, focusing on DZ twin pairs, analyzing ethnic groups within a population separately and utilizing homogenous populations such as the Finns can improve the statistical power of finding QTL significantly. Also, we conclude from the results of this study that even though genetic effects explain a great proportion of height variance, it is likely that there are tens or even hundreds of genes with small individual effects underlying the genetic architecture of height.

Keywords: height, stature, growth, genetic linkage, quantitative trait locus, COL11A1, dizygotic twins, genome-wide association

Sampo Sammalisto, Search for genetic variants influencing human height
Kansanterveyslaitoksen julkaisuja, A1/2008, 124 sivua
ISBN 951-740-759-5; 951-740-760-1 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

TIIVISTELMÄ

Ihmisen kasvu ja aikuisiän pituus ovat tyypillisiä monitekijäisiä ominaisuuksia, joihin vaikuttavat sekä geneettiset tekijät että ympäristötekijät. Geneettisten tekijöiden merkitys pituuden määräytymisessä on huomattava ja on arvioitu, että kehittyneissä maissa nämä selittävät yli 80 % ihmisten välisistä pituuseroista. Vaikka pituus on epidemiologisissa tutkimuksissa assosioitu lukuisiin sairauksiin sekä sosioekonomisiin, psykososiaalisiin ja terveydentilaa kuvaaviin muuttujiin, tiedämme varsin vähän pituuden geenitaustasta.

Tässä tutkimuksessa pyrittiin paikantamaan pituuteen vaikuttavia kromosomialueita genominlaajuisen kytkentäanalyysin avulla sekä tunnistamaan assosiaatioanalyysin avulla näillä alueilla sijaitsevia geenimuotoja, jotka vaikuttavat ihmisten välisiin pituuseroihin.

Ensimmäisessä osatyössä tutkittiin suomalaisesta 277 perheestä koostuvaa perheaineistoa (1417 tutkimushenkilöä) ja merkittävin kytkentälöydös havaittiin kromosomialueella 1p21 (LOD=4,25). Lisäksi havaittiin useita kromosomialueita, (4q35, 9p24, 13q12, 18q21 ja 22q13) jotka saattavat sisältää pituuteen vaikuttavia geenejä (LOD≥2,0). Kromosomialueen 1p21 jatkotutkimukset osoittivat alueella sijaitsevan COL11A1-geenin tietyn geenimuodon assosioituvan pituuteen. (p=0,003). Tämä assosiaatio toistettiin laajassa, 6542 yksilön suomalaisessa väestöaineistossa, jossa tämän geenimuodon suhteen samanperintäiset miehet olivat 1,1 cm ja naiset 0,6 cm pidempiä verrokkeihin nähden. Tämän väestöaineiston avulla arvioitiin, että tunnistettu geenimuoto selittää 0,1 % pituuden vaihtelusta suomalaisväestössä.

Toisessa osatyössä tutkittiin 3817 GenomEUtwin-konsortion (www.genomeutwin.org) Australiasta ja Euroopasta keräämää kaksosperhettä. Koska erimunaiset kaksoset jakavat lähes kaikki kasvuun vaikuttavat ympäristötekijät, tämä asetelma vähentää perheen sisäistä ympäristötekijöistä johtuvaa vaihtelua, joka puolestaan lisää kytkentäanalyysin tilastollista voimaa paikantaa pituuteen vaikuttavia kromosomialueita. Tässä aineistossa merkittävin kytkentälöydös havaittiin kromosomialueella 8q21 (LOD=3,28), jonka lisäksi paikansimme useita kromosomialueita (7p22, 20p13, 21q21 ja Xq25), jotka saattavat

sisältää pituuteen vaikuttavia geenejä (LOD ≥ 2.0). Kromosomialueen 8q21-q24 merkitys pituuden määräytymisessä osoitettiin (p=9x10-8) myös toisessa, 1,552 samanmunaisen kaksosparin genominlaajuisessa assosiaatiokartoituksessa, jossa genotyypitettiin kultakin yksilöltä 317,000 yhden nukleotidin polymorfiaa.

Kolmannessa osatyössä tutkittiin 3032 afro- ja eurooppalais-amerikkalaista perhettä. Tässä aineistossa havaitsimme useita pituuteen kytkeytyneitä geenialueita (11q23, 12q12, 15q25, 15q26, 18q23 ja 19q13), joista merkittävin löydös oli kromosomialueella 15q25 (LOD = 3,0). Kromosomialueen 12q12 kytkentälöydös on myös erityisen mielenkiintoinen, sillä tällä alueella sijaitseva HMGA2-geeni on hiljattain osoitettu assosioituvan pituuteen useissa väestöaineistoissa. Myös kromosomialueen 18q23 kytkentälöydös on varsin kiinnostava, sillä tämä löydös tukee ensimmäisessä osatyössä havaittua kytkentälöydöstä.

Tässä tutkimuksessa paikansimme ja tunnistimme yhden pituuteen vaikuttavan geenimuodon COL11A1-geenissä, joka selittää 0,1 % pituuden vaihtelusta suomalaisväestössä. Myös kromosomialueen 8q21-q24 merkitys pituuden määräytymisessä osoitettiin kahdessa toisistaan riippumattomassa osatutkimuksessa, joista ensimmäinen perustui genominlaajuiseen kytkentäanalyysiin ja toinen genominlaajuiseen assosiaatioanalyysiin. Lisäksi tutkimuksessa osoitettiin, että erityiset aineiston valikointitavat kuten sukupuolten, kaksosparien ja etnisten ryhmien erillisanalyysit voivat lisätä kytkentäanalyysin voimaa merkittävästi. Tutkimuksen merkittävin geenilöydös COL11A1-geenissä antaa myös viitteitä siitä, että vaikka suurin osa pituuden vaihtelusta on geneettisten tekijöiden määräämää, on todennäköistä, että tällaisia tekijöitä on hyvin suuri määrä ja kunkin yksittäinen geenimuodon vaikutus yksilön pituuteen on erittäin pieni.

Avansanat: pituus, kasvu, geneettinen kytkentä, COL11A1, erimunaiset kaksoset, genominlaajuinen assosiaatio

CONTENTS

ABBREVIATIONS

ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
bp	Base pair
CVD	Cardiovascular disease
cM	Centimorgan
DNA	Deoxyribonucleic acid
DZ	Dizygotic
H^2	Heritability (broad sense)
h^2	Heritability (narrow sense)
HWE	Hardy-Weinberg equilibrium
IBD	Identity-by-descent
IBS	Identity-by-state
LOD	Logarithm of odds
Mbp	Mega base pair
ML	Maximum likelihood
mRNA	Messenger RNA
MZ	Monozygotic
QTL	Quantitative trait locus
QTN	Quantitative trait nucleotide
r_A	Coefficient of relatedness
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
V_A	Additive genetic variance
V_C	Variance components

V_D	Dominance genetic variance
V_E	Environmental variance
V_G	Genetic variance
V_I	Epistatic genetic variance
V_P	Phenotypic variance

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by their Roman numerals:

I **Sammalisto S**, Hiekkalinna T, Suviolahti E, Sood K, Metzidis A, Pajukanta P, Lilja HE, Soro-Paavonen A, Taskinen MR, Tuomi T, Almgren P, Orho-Melander M, Groop L, Peltonen L, Perola M. (2005) A male-specific quantitative trait locus on 1p21 controlling human stature. J Med Genet. 2005 Dec;42(12):932-9.

II Perola M, **Sammalisto S**, Hiekkalinna T, Martin NG, Visscher PM, Montgomery GW, Benyamin B, Harris JR, Boomsma D, Willemsen G, Hottenga J, Christensen K, Kyvik KO, Sørensen TIA, Pedersen NL, Magnusson PKE, Spector TD, Widen E, Silventoinen K, Kaprio J, Palotie A, Peltonen L, GenomEUtwin Project. (2007) Combined Genome Scans for Body Stature in 6,602 European Twins: Evidence for Common Caucasian Loci. PLoS Genetics Vol. 3, No. 6, e97 doi:10.1371/journal.pgen.0030097

III **Sammalisto S**, Hiekkalinna T, Doria A, Schwander K, Kelly JA, Bruner GR, Harley JB, Redline S, Larkin EK, Patel, SR, Ewan AJH, Weber JL, Perola M, Peltonen L. Genome-wide Quantitative Trait Locus Screen for Adult Body Height and Body-mass Index in a Combined Sample of 3.032 Families - Evidence for Sex- and Population-specific Genetic Effects. (Submitted)

Also, unpublished data from two other studies are presented.

1 INTRODUCTION

The heritable nature of human height has been recognized for more than one hundred years. It was also through the study of height that the theory for quantitative genetic variation was spawned which provided important advancements in the field of general statistics and gave rise to the mathematical framework for quantitative genetics.

Human height is considered a classical polygenic trait that is determined by the joint action of several genes each with a small individual effect as well as multiple environmental influences. Height is a fundamental human trait which is perturbed in many rare genetic diseases delineating the relevance of growth as an elementary biological mechanism. Human height has been the target of extensive research across many fields of science due to its importance to the individual to and its usefulness as a societal surrogate.

Height can be used as an indicator of early experiences of an individual such as fetal and childhood living conditions and disturbances in normal growth patterns of a child are amongst the most common reasons of referral to paediatric clinics. Growth in infancy is also a well-established measure of the socioeconomic state of a population comparable with gross national product and child mortality. An individual's height is also significantly associated with several health measures, reproductive success, educational attainment and social position. Yet, despite the importance of height as a fundamental human characteristic, little is known about the genetic architechture of height and the specific genes that control growth and attainment of adult height.

Now, with the advent of vast numbers of molecular markers and rapid affordable technologies enabling their genome-wide genotyping as well as the development of powerful computers allows the statistical analysis of large numbers of individuals. Together these advancements should enable us to decipher the specific genetic variants that contribute to the heritable genetic component and quantitative variation observable in the height of human populations. This study aims to integrate the theoretical models developed almost a hundred years ago and state-of-the-art technological achievements with the unltimate goal of elucidating the genetic architecture of human height.

2 REVIEW OF THE LITERATURE

2.1 Height as a quantitative trait

2.1.1 Biological basis of human growth

Adult human height is the endpoint of the longitudinal growth process that occurs during fetal life, childhood and adolescence. The majority of growth in height results from lengthening of the long bones that occurs via cartilage growth at the epiphyseal plate (Figure 1). This cartilage gradually ossifies to form hard bone resulting in lengthening of the bone. Human growth occurs in spurts and can be divided into infancy, childhood and pubertal growth. The growth velocity differs greatly between these periods being greatest in infancy and very rapid in puberty (Figure 2). Growth ceases in late puberty when all cartilage at the growth plates is replaced by bone. This epiphyseal closure is mainly controlled by sex hormones such as testosterone and estrogen[1].

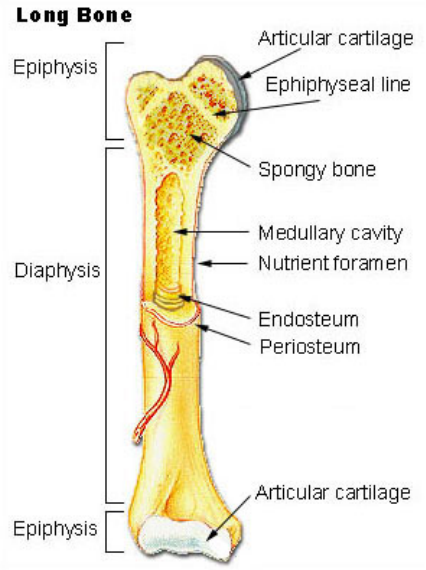

Figure 1. ***Main components of the long bone. Long bones consist of a long shaft (diaphysis) and two articular (joint) surfaces (epiphyses). Growth of the long bones occurs by endochordal ossification at the epiphyseal plate (growth plate). (Source: Wikimedia commons/SEER anatomy and physiology)***

Figure 2. ***Schematic representation of human growth velocity. Stress such as strenuous physical exercise, manual lobor, sleep deprivation, malnutrition or infections at certain life periods may hinder the attainment of full growth potential (Figure courtesy of Dr. Richard Steckel, Ohio State University)***

Growth is mainly controlled by hormones secreted by the endocrine system[2] (Figure 3). Key hormones are estradiol, testosterone, thyroid hormones, growth hormone (somatotropin), insulin-like growth factor-1 (IGF1) and insulin. This hormonal network is complex with several interactions, feedback mechanisms and temporal cues[3]. Sex steroids estradiol and testosterone promote growth in height in childhood and early puberty and accelerate skeletal maturation and epiphyseal closure in late puberty. Both estradiol and testosterone act in association with growth hormone and IGF-1. Thyroid hormones are necessary for normal growth in early infancy and they stimulate osteoblast maturation and accelerate ossification by increasing the levels of growth hormone and IGF-1. Growth hormone is perhaps the single most critical determinant of linear growth and it mainly acts by increasing the proliferation of chondrocytes directly and in concert with IGF-1. The release of growth hormone from the anterior pituitary gland is regulated by growth hormone releasing hormone (GHRH) and somatotropin release-inhibiting hormone (SRIH) but also by the brain neurotransmitters acetylcholine, serotonin and dopamine as well as circulating levels of glucose, IGF-1, estrogens and testosterone[4].

Growth hormone increases the production of insulin, IGF-1 and sex steroids which in turn reduce growth hormone release by a negative feedback system. IGF-1 has similar metabolic effects as insulin but has a more pronounced role in proliferation and maturation of many cell types including chodroblasts and osteoblasts. The effect of insulin on growth is mediated by its main function of ensuring that tissues receive the metabolic fuel needed for growth but also via stimulation of growth hormone and IGF-1 action.

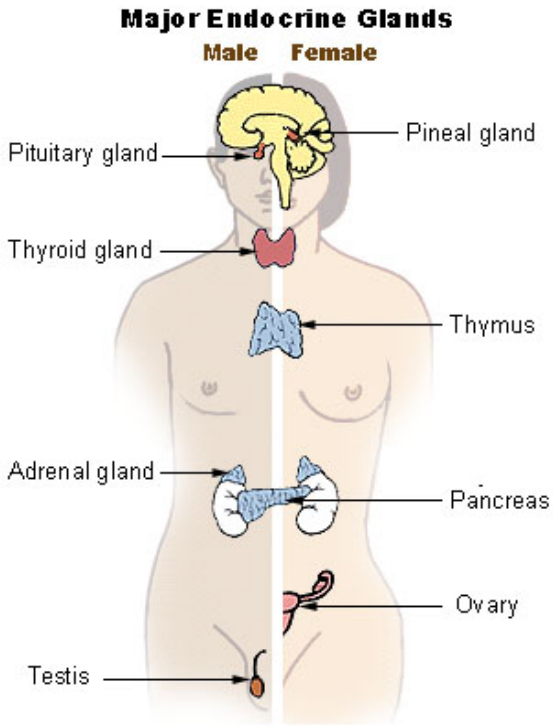

Figure 3. ***Human endocrine system. The endocrine glands secrete hormones that control growth, development, physiological and psychological functions. Most growth-related hormones are secreted by the pituitary gland and gonads with the important expection of IGF-1 which is secreted by the liver. (Source: Wikimedia Commons/SEER anatomy and physiology)***

2.1.2 Genetic background

The study of height has a long standing tradition in genetics that dates back to the work of Sir Francis Galton (1822-1911), a half-cousin of Charles Darwin, who was the first to publish data on the relationship between parent and offspring height 5,6 (Figure 4). This interest in height was probably due to the easy observation even by laymen that in general tall parents usually have tall children and short parent have short children. Also, height has been easy and unambiguous to measure ever since the introduction of standardized measurement scales such as the metric and imperial systems.

Many lines of evidence such as twin, adoption and family studies have established the role of genetic components in the determination of height and subsequently adult height is considered as one of the most heritable quantitative traits known in man[8,9]. It is fair to assume that human growth is a polygenic process where many genes and environmental factors are involved probably also interacting with each other. However, though much is known regarding the biological basis (e.g. hormonal regulation) of human growth, the specific underlying genes that produce the observed variation in human height remain elusive despite extensive efforts to uncover them.

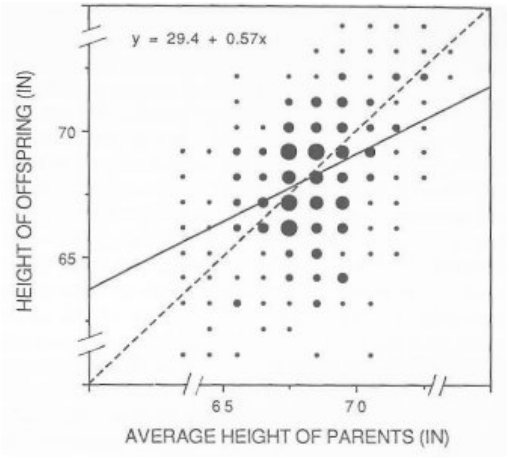

Figure 4. ***The data of Sir Francis Galton[5] showing the relationship between offspring height (928 individuals) as a function of mean parent height (205 sets of parents). The size of the solid circles indicates the number of observations in each category, the broken line perfect correlation and the solid line the linear regression of the observed data. Heritability of height in this sample is equal to the slope of the regression line, 0.57. (After Arnold[7] Source: Wikipedia)***

2.1.3 Environmental influences

Growth potential and environment

Environmental factors play a crucial role in growth and attainment of final adult height. Stunting growth is a universal mechanism of adaptation to poor environmental conditions. The role of environmental factors in the determination of human height is evidenced by 1) analysis of historical height and living condition data and 2) intervention studies of developing countries where living conditions can be improved by supplementary food, nutrients and medicines.

Accumulating data from multiple lines of evidence suggest that suitable environmental conditions allow an individual to reach his or her full genetic growth potential - simplistically speaking, genes set the upper limit to height but whether an individual reaches that height is determined by life conditions. It is clear from longitudinal analyses of body height within populations that increases in the standard of living increases average height - this is called the secular trend (or change) in height. The secular trend has been most obvious in European countries in the last 100 years (Figure 5), where children have gradually matured earlier and reached greater heights on average (reviewed extensively by Cole[10]). This increase in height cannot be attributed to a single factor but is likely due to factors such as improved nutrition, reduction in infectious disease due to better vaccination programs and sanitation as well as the widespread availability of quality health care.

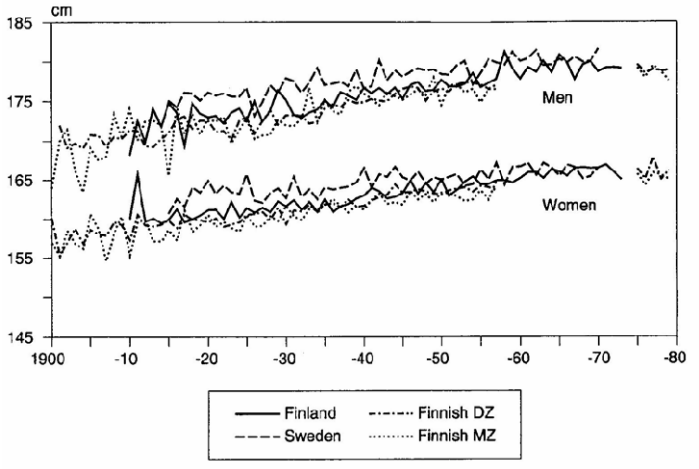

Figure 5. ***Secular trend in mean height of men and women by year of birth in representative sets in Finland and Sweden and in Finnish MZ and DZ twins. Reproduced from Silventoinen 2000[11] with permission.***

Nutrition

Nutrition is a crucial determinant in growth since cell division requires sufficient nutrition to produce necessary proteins and other macromolecules. The growth period in infancy seems most sensitive to nutritional state; this is supported by the observations that most of the differences in growth velocity and absolute height difference between developing and developed countries occurs during the period from 6 months to 24 months of age[12,13] (reviewed also extensively by Cole [14] and Karlberg et al.[15]).

Protein is probably the most important nutrient affecting growth since humans require some amino acids (called essential amino acids) from food that they are unable to produce themselves. Protein deficiency is probably the most important reason for stunted growth in developing countries[16,17].

Minerals and vitamins play an important role in growth as well. Studies have implicated most mineral and vitamin deficiencies as detrimental to normal growth but it is likely that calcium[18], iron[19], vitamin D[20] and vitamin A[21] are the most crucial. Their importance refects their biological role in growth; Vitamin D is essential for calcium absorption which in turn is necessary for bone mineralization, while vitamin A and iron probably influence growth via growth hormone and stress responses [22-25].

Fetal conditions

During pregnancy, growth velocity is very rapid and birth size may also affect adult height[26]. The developing fetus is highly sensitive to environmental factors such as nutrition and toxins such as maternal smoking and alcohol or narcotic use. There are contradicting reports regarding the association of undernutrition during pregnancy and the final height of children, maybe due to catch-up growth in childhood or adolescence[26-28]. There is clear evidence however, that maternal smoking, alcohol and narcotics use reduces both birth weight and the subsequent growth of the child[29-31].

Infectious and chronic disease

Several infectious diseases have been associated with disruption of growth; however most of them are likely to affect growth via reduction of nutrient intake and absorption, loss of nutrients or adverse metabolic processes[32]. Diarrhea caused by food- and water-borne pathogens is by far the most prevalent group of infectious disease stunting growth[33] but also chronic inflammatory diseases such as asthma and type I diabetes have been associated with poor growth[34,35].

Sexual dimorphism

Height is a classical example of a sexually dimorphic trait where males are taller than females in all human populations studied[36] as is evident from Table 1. The average male-to-female ratio in height is 1.07 although there is considerable variation between populations[37,38]. This difference in height is mostly explained by men's longer legs, given that the torso heights are not greatly different[39]. The origin of the height difference between males and females is most commonly hypothesized to result from the sexual dimorphism in 1) hormonal environment and 2) sex chromosome composition.

Sex steroids such as estradiol and testosterone are highly relevant for growth in closure of growth plates in the long bones and they also affect the secretion of other growth-related hormones such as growth hormone and insulin-like growth factor 1[40]. Therefore it is reasonable to hypothesize that the differential sex steroid patterns may produce at least some part of the sex difference in height.

The influences of sex chromosomes are suggested by aneuploidies of the sex chromosomes such as Turner's, Klinefelter's and XYY syndromes but the mechanisms of action are not fully understood. Females with Turner's syndrome lack one copy of the X-chromosome (45, X0) and are characterized by short stature and ovarian failure, while males with Klinefelter's syndrome carry an extra copy (47, XXY) and manifest mild mental retardation and are slightly taller on average compared to males with the normal karyotype. XYY males (47, XYY) males are taller than average but are devoid of other clear characteristic phenotypic manifestations although some reports show evidence for increased learning disabilities and possibly slightly impaired intelligence. Taken together, the stature manifestations in these aneuploidies suggest that sex chromosomes may influence height via dosage effect of pseudoautosomal and Y-specific growth genes although the specific genes have yet to be identified[41-44].

2.1.4 Associations with health

Height has been to shown to positively correlate with several health indicators and overall mortality[45,46] and on average short people have poorer health than tall people. However, most of these associations may result from the association of growth and final adult height to childhood living conditions which in turn also associate with health measures and outcomes[47-49] i.e the confounding role of childhood living conditions (Figure 6).

The inverse correlation of height and incidence of cardiovascular diseases (CVD) such as myocardial infarction, coronary heart disease and stroke is perhaps the most well studied association between height and health[50-53]. Although most investigators attribute these associations to childhood living conditions some have proposed shared etiology, such as the positive correlation of height and blood vessel diameter[54].

Numerous associations have also been reported between tall height and cancers such as breast and prostate cancer[55-57]. Many investigators have attributed these positive associations between height and cancer as suggestive that high childhood calorie intake which promotes growth in stature may also increase cancer risk[58,59].

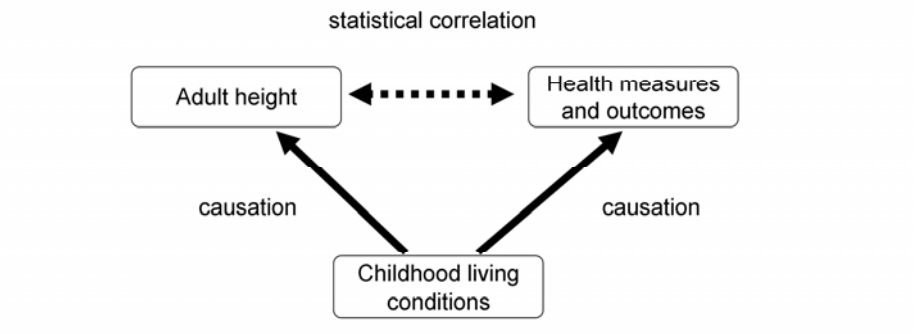

Figure 6. ***Confounding model for association between height and health measures. Childhood living conditions (confounder) which affect adult height and health measures and outcomes causes spurious associations between the two.***

2.1.5 Socioeconomic and psychosocial relevance

Height is a fundamental human characterististic that has several social and psychological associations. Height correlates with education level and socioeconomic status in most human societies; tall people are usually more educated and have better socioeconomic status than short people[60-62]. These correlations can be attributed to two factors 1) parents of higher socioeconomic status are able to provide better childhood living conditions for their children, which in turn promotes the full attainment of growth potential and 2) the education level and socioeconomic status of parents predicts the education level of their children well. Therefore, this correlation may be due to confounding of a favorable childhood environment and transmission of socioeconomic status[62]. However, height is also correlated with social mobility (shift in socioeconomic position between child- and adulthood)[63-65].

It has been shown in psychological studies that humans hold an unconscious association of status, power and height; we tend to picture people of authority as physically big[66,67]. For this reason it may be easier for tall persons to excel in social hierarchies since they possess inherent status due to their physical appearance.

Height is also positively correlated with reproductive success in contemporary societies especially in men mainly because taller than average men are found to be more attractive to women[68-70]. Taller men enjoy more socioeconomic success than shorter men, which may also attract partners, however height has also been shown to be an independent factor for reproductive success[69,71]. Interestingly there is some data indicating that women that are shorter than average enjoy greater reproductive success also suggesting that the favoring of tall men by short women maintained the sexual dimorphism observed in stature[72]. However, also the height of an individual is an important factor in mate choice and positive assortative mating for height is well established and the height of spouses correlates well[73,74]. Assortative mating refers to the phenomenon where individuals choose their mate based on some specific phenotypic trait that they are similar (positive assortative mating) or dissimilar (negative assortative mating) for.

2.1.6 Population differences in height distribution

Contemporary human populations vary tremendously in height: for example in the Netherlands the average height of males is 183.1 cm and females is 170.0.cm[75] while the Ituri (Pygmy) males are on average 144.4 cm and females 136 cm tall[76]. However, in Western Europe and the United States there is much less variation in average heights probably because the general environment is very similar in respect to height (Table 1)

Table 1. ***Average height in selected countries by sex. Note, that mean height in Cavalaars 2000 is age standardized to eliminate cohorts and shrinkage effects so these data are not fully comparable to the data from Australia and the US.***

Country	Age Group	Males	Females	Reference
Australia	24-44	176.3	162.9	Australian Bureau of Statistics 1995: [77]
Denmark	20-74	177.1	165.2	Cavelaars 2000[60]
Finland	20-74	176.6	163.5	Cavelaars 2000[60]
France	20-74	173.1	161.8	Cavelaars 2000[60]
Germany	20-74	175.4	162.8	Cavelaars 2000[60]
Italy	20-74	172.2	162.1	Cavelaars 2000[60]
Netherlands	20-74	178.7	167.1	Cavelaars 2000[60]
Norway	20-74	178.9	165.8	Cavelaars 2000[60]
Spain	20-74	170.0	160.3	Cavelaars 2000[60]
Sweden	20-74	177.9	164.6	Cavelaars 2000[60]
Switzerland	20-74	175.4	164.0	Cavelaars 2000[60]
US blacks	30-39	177.1	163.3	Godoy 2005[78]
US whites	30-39	176.8	163.3	Godoy 2005[78]

Historically, most of the population differences in height and body shape are likely to be evolutionary adaptations to local climate (temperature and humidity) and terrain (forest, plain or mountainous). These adaptations have occurred over hundreds of generations since the genes that govern anthropometric traits change slowly.

Nutrition and living conditions are important sources of height variation especially in contemporary human populations. Migration studies have shown rapid increases in average height in only a few generations[79-81]. Clearly genes can change very little in this time frame and the increase in height must be due to improved diet and reduced environmental stress factors.

2.2 Human genetic variation and the human genome

2.2.1 Organization of the human genome

The haploid human genome consists of roughly 3.2 billion base pairs of double stranded deoxyribonucleic acid (DNA) and is organized in 24 distinct chromosomes (22 autosomes, X- and Y-chromosome) (Figures 7 and 8). Each individual receives one chromosome of each type from each parent so that an individual's genome consists of 22 autosomal chrosome pairs and a sex chromosome pair which may be either XX (female) or XY (male).

The human genome contains roughly 20,000-30,000 genes which are defined as nucleotide sequences that code for ribonucleic acid (RNA) -molecules that have some specific function(s) in the cell (protein-coding, structural or regulatory). Genes take up less than 2% of the total DNA content in the genome while the rest of the DNA remains still largely uncharacterized although an increasing number of regulatory and other functional domains are being identified in intergenic regions. The human genome contains large amounts (>50%) of repetitive sequence elements that may be either in tandem (satellite DNA) or interspersed (e.g. LINE, SINE) repeats that may also be transposable[82]. The function of these repeated sequence elements are unknown but it has been suggested that they are involved in genome evolution by rearranging the genome (reviewed extensively by Kazazian[83]). Human genes are consisted of exons that determine the amino acid sequence of the nascent proteins and introns (intervening sequences) that are located between successive exons.

The genomic lengths of genes, exons and introns vary greatly. The genomic length of the average gene is 27 kb although there is tremendous variation, for example between the insulin (1.4 kb) and dystrophin (2,400 kb) genes. The average number of exons in human genes is 9 but there is tremendous variation e.g. beta-1-adenergic receptor (1) and titin (363). Interestingly, introns are on average (3,365 bp) larger than exons (122 bp), but the lengths of both vary greatly.

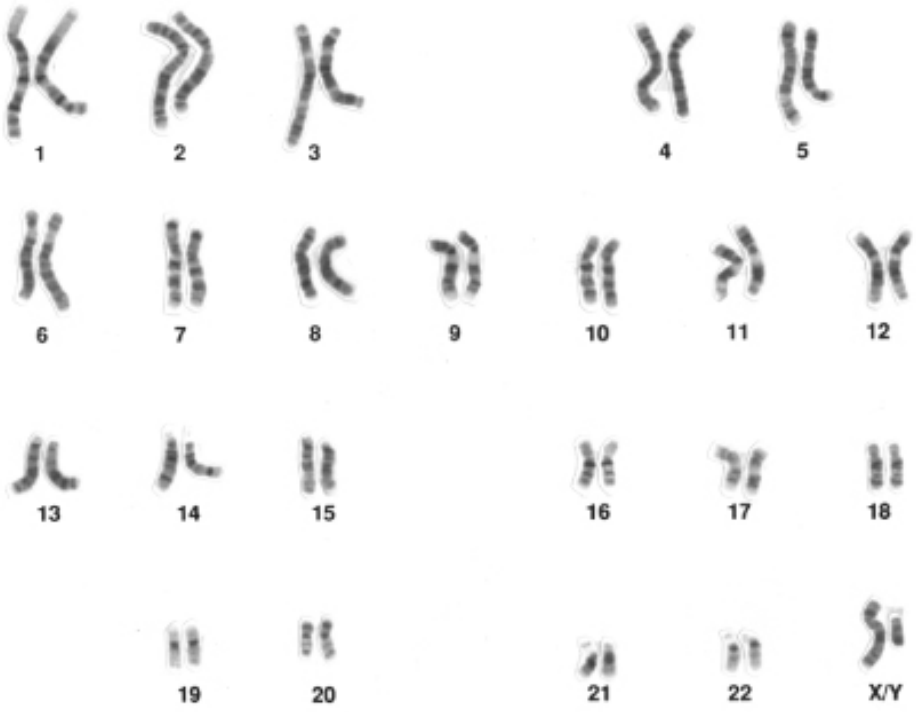

Figure 7. ***Human male karyotype. Karyotype shows each autosomal chromosome pair (1-22) at the sex chromosomes X and Y. Source: National Human Genome Research Institute Talking Glossary of Genetic Terms (www.genome.gov/glossary.cfm)***

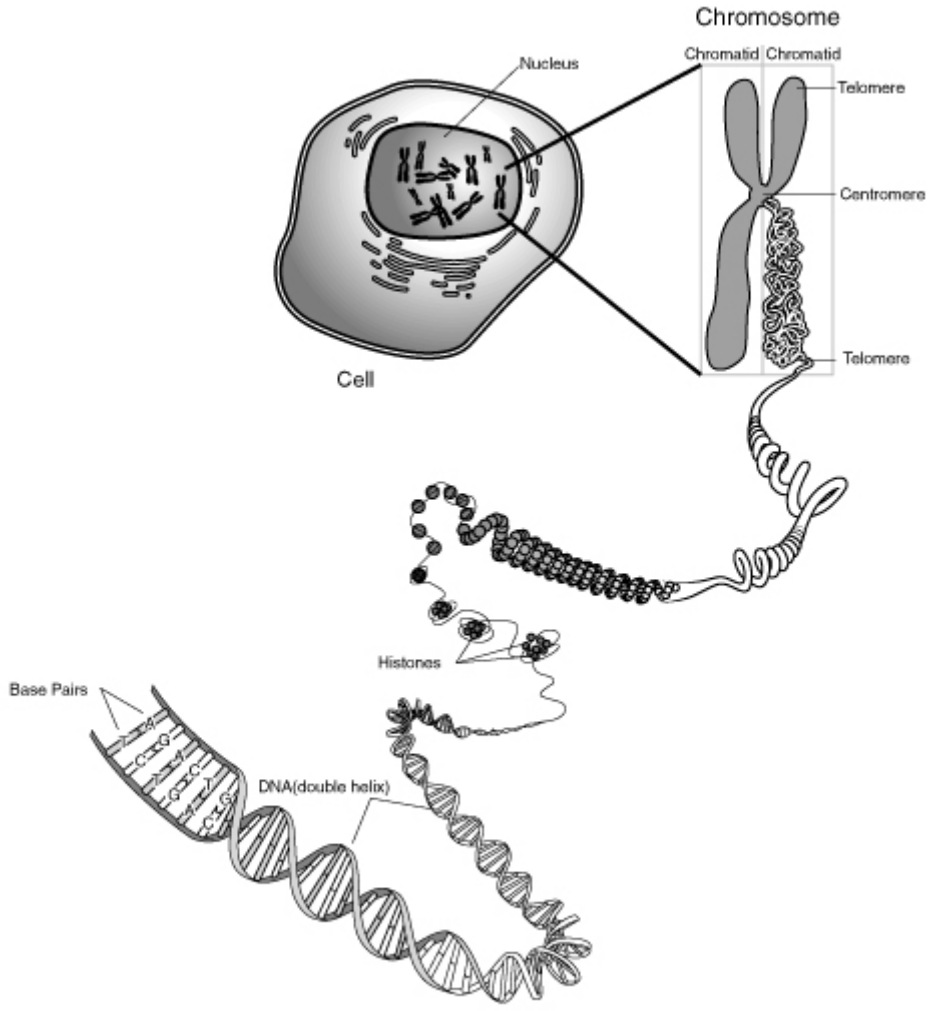

Figure 8. ***DNA and chromosomes. Double stranded DNA is wrapped around histone proteins which are further packed tightly to form the chromosomal structure which are contained in the nucleus of the cell (www.genome.gov/glossary.cfm)***

2.2.2 Emergence of genetic variation

Most of the genetic variation in humans arises from 1) mutation that produces novel variation in the population, 2) genetic recombination that rearranges that variation between homologous chromosomes and 3) random assortment that shuffles these chromosomes into gametes.

Mutation is the most important force that produces novel variation. Mutation rates in humans are extremely low, approximately 2.5 x 10^{-8} mutations per nucleotide site or 175 mutations per diploid genome per generation[84], due to the highly efficient DNA repair mechanisms that identify and correct mutated sequences introduced by DNA replication errors or mutating agents. Mutations affecting only one or few nucleotides may be either 1) substitutions that replace nucleotides with other types of nucleotides, 2) insertions that add nucleotides to a sequence or 3) deletions that remove nucleotides from a sequence (Figure 9).

Figure 9. ***Point mutations. (A) Substitution, where one nucleotide is changed to another. (B) Deletion, where a nucleotide is lost from the sequence (C) Insertion, a nucleotide is gained by the sequence.***

Mutations may also involve large portions of chromosomes; these can be divided into 1) deletion, 2) duplication, 3) inversion or 4) insertion of a given chromosomal region or 5) translocation where chromosomal regions are interchanged between nonhomologous chromosomal regions (Figure 10).

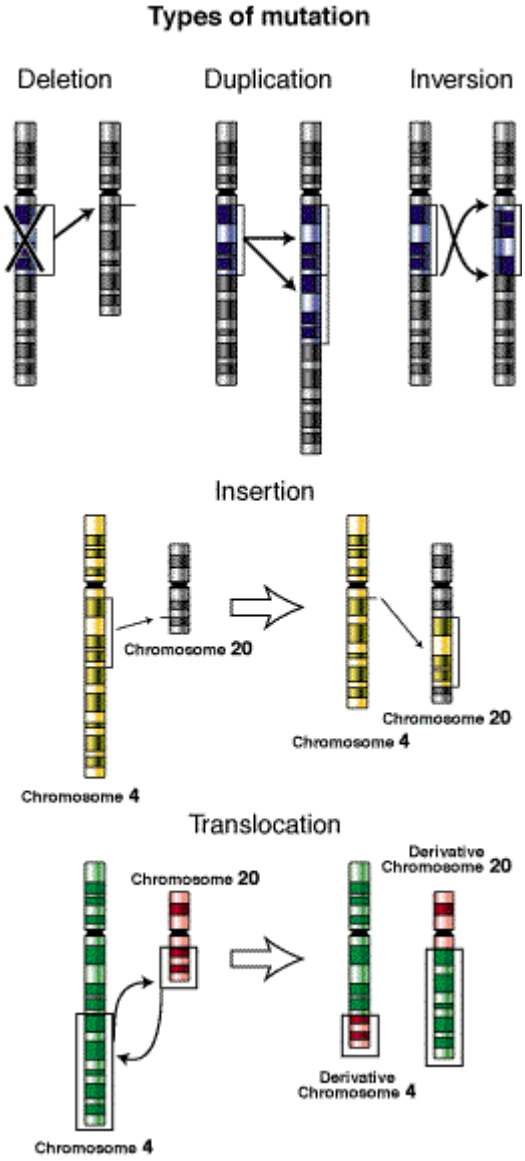

Figure 10. ***Chromosomal rearrangements.Mutations may also involve large segments of chromosomes and these rearrangements can be divided into deletions, duplication, inversions, insertions and translocations (Source: www.genome.gov/glossary.cfm).***

Genetic recombination is a phenomenon where homologous chromosomes (i.e. maternally and paternally inherited copies of a given chromosome pair) interchange a certain segment (Figure 11). Recombination occurs in meiosis, which is the process in which gametes are produced from germline cells, and results in recombinant chromosomes that contain different combinations of alleles (haplotypes) than the parent chromosomes. Recombination is a common phenomenon and the average rate of genetic recombination in humans is 1.13 per 100 Mbp[85], so on average there are 36 recombination events per meiosis per genome.

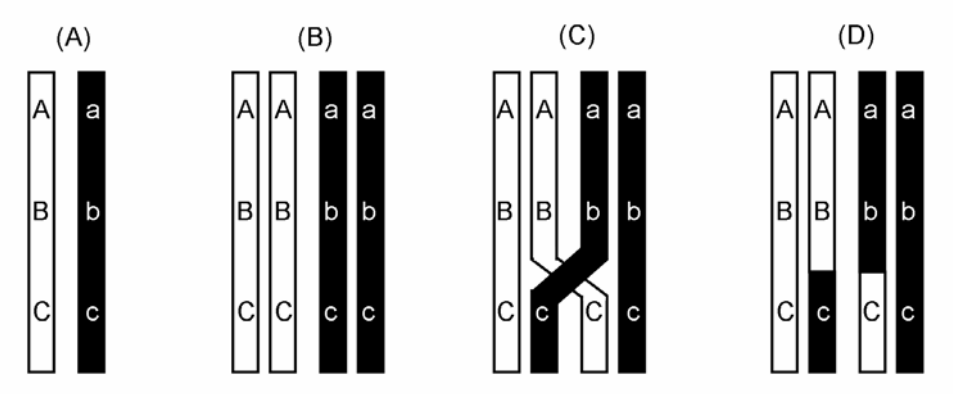

Figure 11. ***Schematic representation of genetic recombination in meiosis between homologous chromosomes. (A) Parental chromosomes of an individual that is heterozygous for three loci and has haplotypes ABC and abc.The open rectangle is the maternal and the solid rectangle is the paternal copy of the chromosome pair. (B) Chromosome duplication. (C) Formation of chismata and crossing over where maternal and paternal chromosomes exchange genetic material (D) forming two recombinant chromosomes that carry novel haplotypes ABc and abC. Figure courtesy of Tero Hiekkalinna.***

Independent assortment in meiosis was already described by Gregor Mendel and it refers to the combinations maternal and paternal chromosomes that may be sampled to a single gamete in meiosis. For each chromosome pair either the maternal or the paternal chromosome may be sampled to a haploid gamete; therefore since the assortment of chromosomes is independent of each other between nonhomologous chromosomes and there are two copies of each homologous chromosome there are $2^{23} \approx 8.4$ million possible combinations of maternal and paternal chromosomes that the gamete may carry (Figure 12).

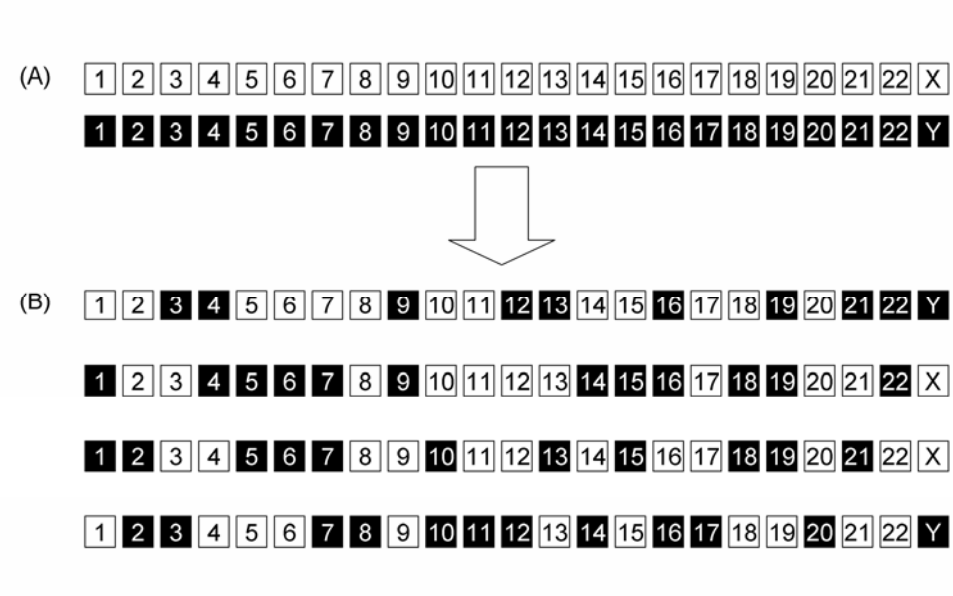

Figure 12. ***Independent assortment in meiosis. (A) Chromosomes of a given individual, maternally inherited choromosomes with open and paternally inherited chromosomes with solid squares, (B) Four possible combinations of maternal and paternal gametes out of the $2^{23} \approx 8.4$ million possible combinations. This diagram ignores genetic recombination which adds a further level of diversity by the mixing of maternal and paternal genetic material within homologous chromosomes.***

2.2.3 Genetic variants used in genetic mapping

Approximately 0.1% of the human genome is variable at an appreciable level (e.g. polymorphism where at least 1% of the population possesses a less common variant) and this genetic variation can be divided into two main categories: 1) sequence polymorphisms where a single or few nucleotides differ between some individuals or 2) length polymorphisms where there is variation between the lengths of the genomes of individuals due to insertion, deletion or copy number variation of certain sequence elements. In the context of genetic mapping there are two types of genetic variants that are most relevant: 1) single nucleotide polymorphisms (SNP) and 2) short tandem repeats (STR) that are also commonly referred to as microsatellites (Figure 13).

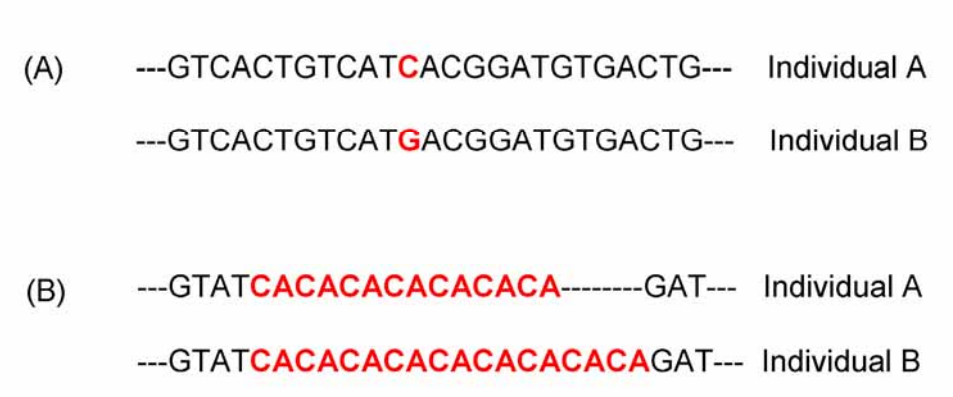

Figure 13. ***Genetic variants commonly utilized in genetic mapping. (A) Single nucleotide polymorphism. At this location there is variation in the nucleotide content of individuals; some have a cytocine C nucleotide such as individual A does, while others, like individual B, have a guanine (G) residue at this position. (B) Microsatellite sequence. At certain positions there are tandemly repeated sequences where individuals may differ in the number of copies of the repeat sequence. For example for a CA-dinucleotide repeat, some individuals may have seven copies of that repeat as individual A does and some may have nine copies, like individual B.***

SNPs are historical point mutations that have persisted over the course of evolution either because they have had some beneficial effect in some environment at some point in time and have been a target for natural selection or because they have been selectively neutral and have persisted simply due to stochastic factors (chance). SNPs have accumulated in the genome over the course of evolution and are very frequent, current estimates are that there are more than 10 million SNPs in the human genome that have a minor allele frequency of 1% in the studied populations (dbSNP[86] build 126). Most SNPs are bi-allelic due to the low probability of mutation occurring at the exact same nucleotide position - also even if mutation were to occur at the same location it is unlikely that it would be evolutionarily preserved and observable today.

Microsatellites (STRs) are loci that consist of tandemly repeating units of 1-4 base pairs (bp) in length. Sequence repeats are highly prone to DNA replication errors due to a phenomenon called replication slippage and tend to be duplicated in tandem. This is the reason why the mutation rates of microsatellite loci are substantially higher compared to other loci and they are highly polymorphic with up to 10 alleles per locus. There are over 300,000 known microsatellites in the human genome (NCBI Build 36.1).

Another class of human genome variation that has received attention recently are copy number variants (CNVs). CNVs are large genomic regions ranging from a few kb to several Mb that vary in copy number between individuals (average 250 kb). CNVs are common in the genome and it has been estimated that 12% of the human genome consists of CNV sites[87]. Although CNVs are rarely used in genetic mapping many investigators have suggested that they may be important sources of human phenotypic variation because they result in gene dosage differences between individuals[88]. CNVs may also be important structural elements in the genome by providing a structural basis for genetic recombination and chromosomal rearrangements[89]. To date, there are over 8000 known copy number variable regions in the Database of Genomic Variants (http://projects.tcag.ca/variation/)

2.2.4 Human population genetics

Population genetics focuses on the nature and sources of genetic variation and predicting changes in the relative frequencies of different genotypes in a population. A population in the genetic sense is defined as a group of *interbreeding* individuals that are able to produce viable offspring. In addition to biological compatibility interbreeding also requires geographical proximity and absence of other potential limitations for breeding such as social or cultural barriers. Since especially in humans many of these limitations are present (perhaps even more so in the past), there is great phenotypic diversity observable between contemporary human populations.

Evolution is in its basic form driven by production of genetic diversity (via mutation, random assortment, genetic recombination and sexual mating) and *natural selection* of genotypes that are favorable in the population's current environment. However, stochastic factors such as *genetic drift*, *the founder effect* and *population bottlenecks* are important as well. Genetic drift refers to random change in allele frequencies, which results from the random sampling of gametes from one generation to another (Figure 14). The founder effect is another form of genetic drift that refers to the establishment of a new population by a small number of individuals that may carry only a fraction of genetic variation of the original source population. Due to extensive migration in human evolutionary history founder effects are rather common. Population bottlenecks are also a form of genetic drift; they are evolutionary events where a significant portion of the population dies or is otherwise prevented from reproducing and these are also fairly common in human history due to famines, wars and disease epidemics.

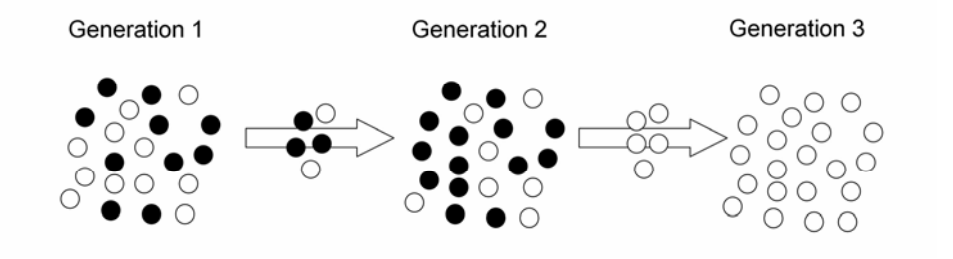

Figure 14. ***Genetic drift. In the first generation both alleles (open and solid circles) of a bi-allelic locus are equally frequent in a small population. However, due to random sampling of gametes in sexual mating the solid allele is slightly over-represented in the gamete pool and thus allele frequencies are not equal in generation two. After another round of gamete sampling, the solid alleles are not sampled to the gamete pool at all and it is lost completely from the population in generation three where only open alleles are present.***

2.3 Quantitative genetics and genetic mapping

2.3.1 Quantitative variation of human traits

In human genetics emphasis has gradually been shifting in recent years from analysis of discrete, usually dichotomous (yes/no) phenotypes to continuous or quantitative phenotypes which can attain any value within a certain range (or in some cases specific values such as positive integers). This emphasis is largely due to the realization that even though often the phenotype of medical interest is based on clinical criteria (such as the manifestation of certain disease), the risk factors for that disease may be quantitative in nature (hormone, enzyme or macromolecule levels, blood pressure or mRNA levels). Unless there is specific justification for dichotomizing a continous variable (e.g. bimodal distribution) utilizing the full quantitative information results in better statistical power to detect underlying genes[90,91].

Quantitative genetics is by no means a new field; it has a well-established track record in plant and animal breeding but human geneticists are often unfamiliar with the history and current state of quantitative genetics. This is somewhat surprising since the field of quantitative genetics was born out of studies of human height by Francis Galton, Karl Pearson and Ronald Fisher in the late 19th and early 20th centuries. The work of these scholars also provided tremendous leaps for general statistics methodology introducing techniques such as multiple regression, multiple correlation and variance-components partitioning which is the basis on analysis of

variance (ANOVA). For example correlation was introduced by Galton to describe the relationship between parental and offspring height in his data[6,92]. Galton's and Pearson's work on general inheritance of quantitative characteristics in turn led them to introduce the concepts of multiple correlation and multiple regression[93,94].

2.3.2 Genetic basis of quantitative variation

In the late 19th and early 20th scholars in the field of genetics were divided into two schools of thought: the Mendelian and Galtonian schools. Proponents of Mendel's laws of inheritance believed that phenotypic traits were transmitted in discrete units (genes) while the Galtonian school assumed that phenotypic traits were transmitted directly from parent to offspring (instead of transmitted by genes). In a landmark paper by Fisher in 1918[95] he was able to resolve the dispute and unite these schools of thought by postulating that quantitative variation and the phenotypic correlation between relatives could be explained by a large number of loci segregating according to Mendel's laws each contributing to the phenotype in minute proportions (polygenes) as well as environmental influences producing continuous variation.

Let us consider a simple example illustrating Fisher's model of the polygenic nature of quantitative traits by comparing the phenotypic distributions when the trait value is completely determined by one, two or three bi-allelic genes. In this example we assume that the phenotype of the heterozygotes is exactly intermediate in respect to the two homozygotes (additivity), each gene has an equal effect on the phenotype and that alleles at each locus are equally frequent and genotypes are in Hardy-Weinberg equilibrium. As can be clearly seen in Figure 15, as the number of loci contributing to the trait increases the number of trait value classes increases gradually approaching a normally distributed continuous trait. Also, it is noteworthy that when there are more than one gene controlling the trait the phenotype to genotype relationship becomes less clear; for example if there are two genes controlling the trait genotypes $A_1A_1B_1B_2$ and $A_1A_2B_1B_1$ produce identical phenotypes (trait values).

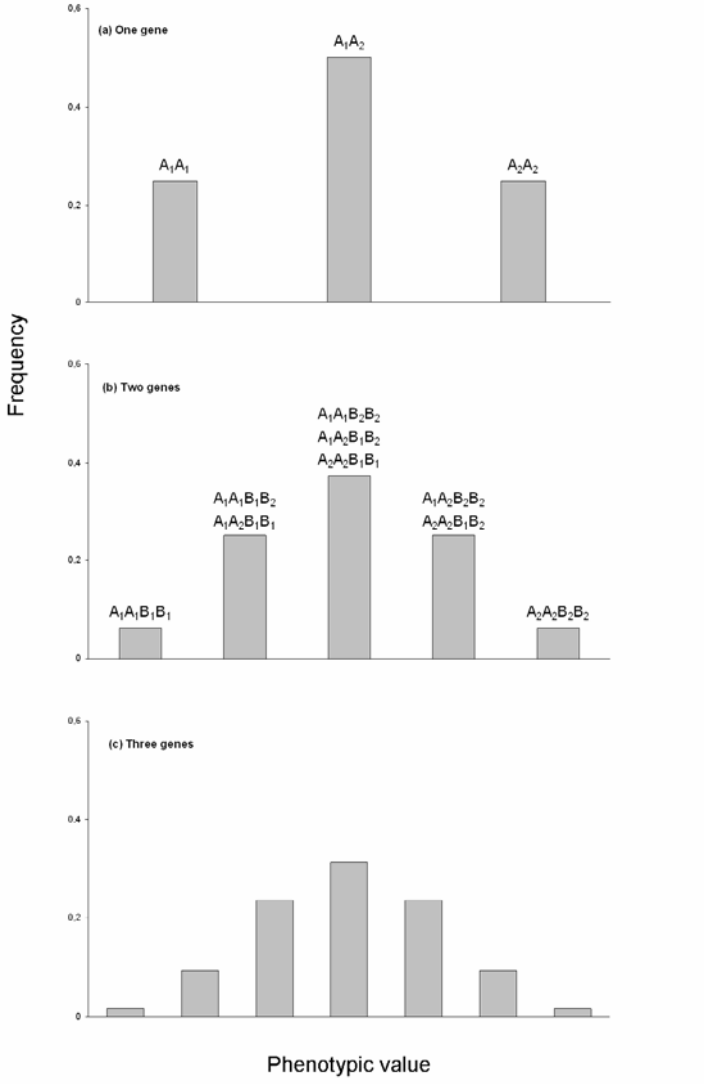

Figure 15. ***The phenotypic values when (a) one, (b) two or (c) three genes determine the trait value. It is assumed that all alleles are equally frequent, the genotypes are in Hardy-Weinberg equilibrium and that there is no dominance at the loci. The genotypes that result in a certain phenotypic value are shown for one and two genes.***

In addition to their polygenic background, quantitative traits are usually also influenced by environmental factors that may act separately or in concert with genes (gene-environment interaction). The varying environments of individuals carrying certain genotypes at a QTL also contribute to the continuous nature of many phenotypes as is shown in Figure 16.

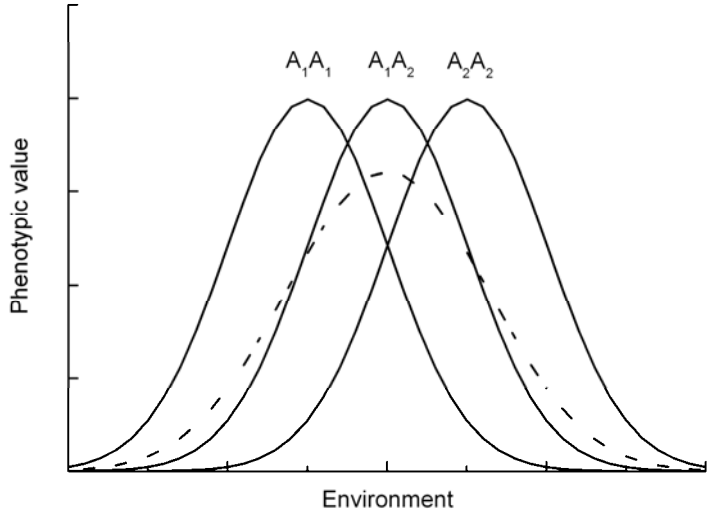

Figure 16. ***The effect of environment on quantitative variation. The solid lines indicate the phenotype value distribution for each genotype of a bi-allelic QTL and the broken line the combined phenotypic distribution of all three genotypes.***

Polygenic inheritance may also relate to dichotomous traits, such as disease states (affected/unaffected). Such traits may partially be determined by an underlying genetic liability that has a continuous distribution and the disease is manifested when this liability exceeds a certain threshold (Figure 17). Relating this to the previous example, the disease may be manifested when a sufficient number of risk alleles (e.g. A_2 and B_2) produce a large enough genetic liability. This theory of the polygenic backroung underlying discrete trait is called the liability-threshold model and was introduced by Pearson and Lee in 1901[96].

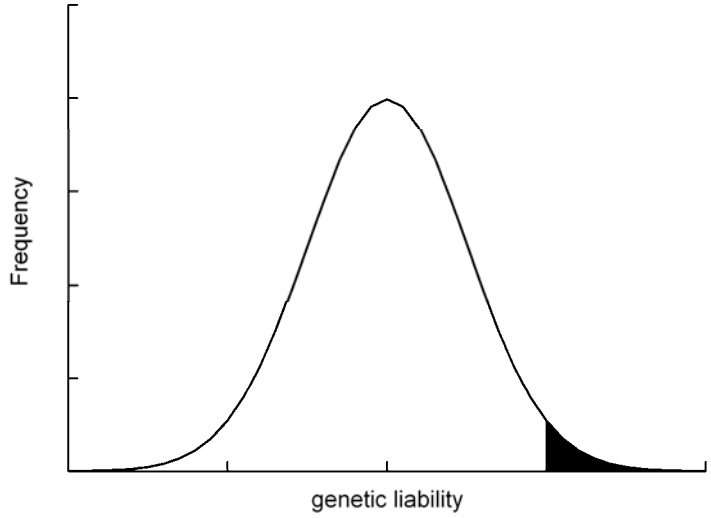

Figure 17. ***The liability-threshold model for dicrete traits. The underlying genetic liability for manifesting a discrete trait such as a disease is continuous in the population and individuals whose liability exceeds a certain threshold (dark shading) manifest that trait.***

2.3.3 Quantitative genetics models

Quantitative genetic models are based on the decomposition of the phenotypic value (P) of an individual to genetic (G) and environmental (E) components. In the simplest model we consider genetic and environmental effects as independent aggregate effects, so that the phenotype can be described as

$$P = G + E$$

Since we are interested in phenotypic and genotypic variability, it is useful to examine quantitative traits and their components in terms of variances so that

$$V_P = V_G + V_E + 2COV_{GE}$$

where V_P, V_G, V_E and COV_{GE} are phenotypic, genetic and environmental variance and COV_{GE} genotype-environment covariance, respectively. The genotype-environment covariance term refers to the direction of the genotype and environment effects (positive when genotypes with higher trait value are also in better environments and genotypes with lower values are in worse environments), not to gene-environment interaction (the dependence of gene action from the environment). However, in many instances (e.g. planned experiments) the covariance term can be ignored because it is close to zero. Then this model is simplified to

$$V_P = V_G + V_E$$

2.3.4 Heritability

Definition of heritability

Heritability is defined as the proportion of phenotypic variance that is explained by genetic factors. The concept of heritability was introduced by Fisher in his seminal 1918 paper[95]. Most commonly in the human genetics literature, heritability refers to the broad sense heritability, H^2, which reflects the proportion of all genetic effects (additive, dominant and epistatic) influencing phenotypic variation.

$$H^2 = \frac{V_G}{V_P}$$

In many instances (e.g. QTL mapping, rate and response to directional selection) we wish to decompose the aggregate genetic effects (V_G) further to their additive (V_A), dominant (V_D) and epistatic components (V_I)[97].

$$V_P = V_A + V_D + V_I + V_E$$

The proportion of phenotypic variance attributable to additive genetic effects is termed the narrow sense heritability, h^2.

$$h^2 = \frac{V_A}{V_P}$$

Properties of heritability estimates

Although the definition of heritability is very clear, there are many common misperceptions in its interpretation. Heritability is always specific to the sample where it is estimated, in other words no trait possesses inherent heritability. It is incorrect to state that a trait is highly heritable on the basis of a high heritability estimate, because heritability is not an attribute of a trait. For example, changing the environment is likely to change the heritability estimate even when genetic variation and all other factors are fixed. Therefore, also the *exact* comparison of heritability estimates across samples is not feasible since environmental factors are very difficult to standardize at least in non-experimental populations.

It should also be noted that a specific heritability value gives absolutely no indication to mode of inheritance, number of loci nor the extent to which the phenotype is "controlled" by genes (since the environment may influence the heritability greatly).

Heritability estimates always relate to a specific *sample* in a specific environment and are therefore not applicable to any given individual. For example, if the heritability of a given disease is 0.60 in some sample this does not mean that for a given individual from that sample 60% of the risk of developing that disease is due to that individual's genetic makeup. One must remember that heritability is a proportion of variances, and since variance is meaningless for a single data point (such as a single individual) also heritability is meaningless for a single individual. In essence heritability is analogous to the arithmetic mean, which is also a property of a given sample not a property of a given trait or a single data point (e.g. an individual).

Estimation of Heritability

The estimation of heritability is based on the correlations of phenotypic values between different relative pair classes that have varying coefficients of relatedness (Table 2). If a trait is controlled by genes we would assume that close relatives resemble each other phenotypically more than distant relatives since they share more of their genes on average. Heritability can be estimated from phenotype data of relative pairs by using correlation and regression methods or analysis of variance (ANOVA).

Table 2. ***Coefficient of relatedness r_A for selected relative pairs. r_A describes the average proportion of shared genes between individuals.***

Relationship	**Coefficient of relatedness r_A**
Monozygotic twins	1
Dizygotic twins	1/2
Parent-offspring	1/2
Full siblings	1/2
Grandparent-grandchild	1/4
Half siblings	1/4
Avuncular	1/4
First cousins	1/8

The classical method of heritability estimation is the *twin method* in which one compares the intraclass correlations of monozygotic (MZ) twins who are genetically identical and dizygotic twins (DZ) who share half their genes on average. The basic idea is that if there are genetic components, there should be less variation among MZ pairs than among DZ pairs. In the classical twin method it is assumed that both types of twins share environmental effects similarly (although this has been criticized) and the heritability is approximately twice the difference in correlation between MZ and DZ twin pairs[98,99].

$$h^2 = 2[corr(MZ) - corr(DZ)]$$

For other types of relatives such as parent-offspring, full sibling and half-sibling pairs heritability can be estimated using regression methods, ANOVA or maximum likelihood (ML) methods based on variance components. In general, heritability can be estimated as the proportion of the regression coefficient b and the coefficient of relationship r_A.

$$h^2 = \frac{b}{r_A}$$

2.3.5 Genetic covariance between relatives

The genetic covariance among relatives is usually described in terms of *identical-by-descent (IBD)* sharing of alleles. By definition, individuals share an allele at a given locus IBD if they have inherited that allele from a common ancestor. In some cases it is also useful to determine another measure for allele sharing between individuals, called *identity-by-state (IBS)*. Two individuals share an allele IBS if they have an identical copy of a certain allele at a given locus. IBS does not consider the ancestral origin of alleles but simply the identity of specific alleles. The difference between these two measures is exemplified in Figure 18.

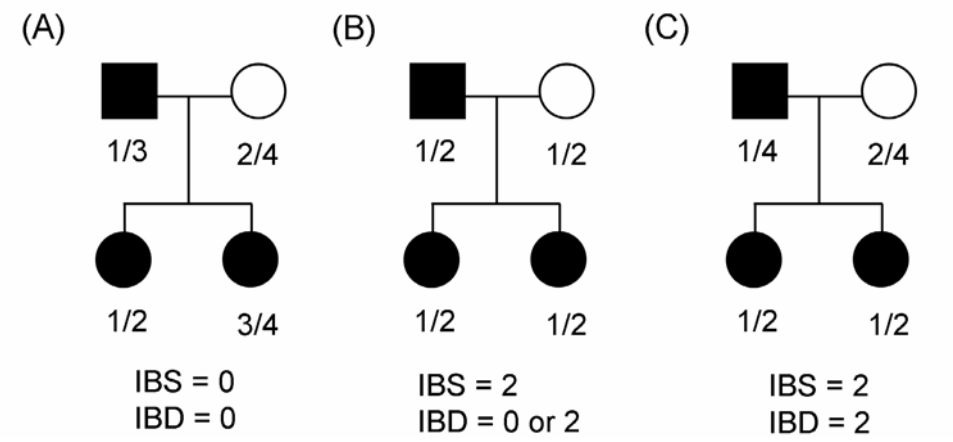

Figure 18. ***IBS and IBD sharing between sib pairs. (A) The sibs do not have any alleles in common, therefore IBS=0. Further, since one sib has inherited allele 1 from her father and 2 from her mother and the other has inherited allele 3 from her father and 4 from the mother, also their IBD=0. (B) Sibs have both alleles in common (i.e. both have allele 1 and allele 2), therefore they share 2 alleles IBS. In this case the parental origin of alleles can not be determined since both sibs may have inherited allele 1 from their father and allele 2 from their mother, or vice versa. If both sibs have inherited allele 1 (and allele 2) from the same parent IBD=2 but if they have inherited allele 1 (and allele 2) from different parents IBD=0. (C). Sibs have both alleles in common (i.e. both have allele 1 and allele 2), therefore they share 2 alleles IBS. Both sibs must have inherited allele 1 from their father and allele 2 from their mother, therefore also IBD=2.***

At autosomal loci two diploid individuals may share either 0, 1 or 2 alleles IBD (or IBS) and we can derive probabilities for each IBD state and the expected proportion of IBD sharing across all loci over the genome by applying Mendel's laws and simple probability calculus. Consider a fully informative situation where both parents are different heterozygote types: the mother has genotype 1/2 and the father genotype 3/4. According to Mendel's laws of segregation and independent assortment each parent produces both types of gametes in equal proportions and the probability of two specific gametes joining to form a gamete is simply a product of their frequencies. Therefore, the probability of each offspring genotype (1/3, 1/4, 2/3 or 2/4) is 0.25 (Table 3).

Table 3. ***Possible offspring genotype for when parental genotypes are 1/2 and 3/4. Parents produce both gamete types in equal proportions and because the joining of gametes is random in fertilization each possible offspring genotype is equally likely.***

		Paternal gametes	
		3	4
Maternal gametes	1	13	14
	2	23	24

Because the genotypes of the offspring are independent of each other, the probability of siblings having a specific genotype pair (e.g. sib A 1/3 and sib B 2/4) is simply the product of the respective genotype frequencies. If we work out all possible genotype pairs, their probabilities and the IBD states we can derive the expected IBD distribution for sib pairs (Table 4). Since we know that all genotype pairs for the siblings are equally likely (the probability for each pair is $0.25^2 = 0.0625$) we can ignore this term and simply calculate the occurrence of each IBD state; therefore the probability that sibs share zero alleles IBD is 4/16 (=0.25), one allele 8/16 (=0.5) and two alleles IBD is 4/16 (=0.25).

Table 4. ***Genotype combinations and IBD states for a sib pair when parental genotypes are 1/2 and 3/4. The genotypes of sibs are independent of each other, therefore each genotype pair is equally likely.***

		Sib B			
		1/3	1/4	2/3	2/4
Sib A	1/3	2	1	1	0
	1/4	1	2	0	1
	2/3	1	0	2	1
	2/4	0	1	1	2

The theoretical IBD distribution for any type of relative pair can be derived in the same manner; however this becomes rather tedious for more distant relative pairs. The expected IBD is also called as *coefficient of relationship* r_A and can also be thought of as the probability that an allele drawn at random from one individual is IBD to one of the alleles of the other individual. The coefficient of relationship is also the average IBD across all loci over the genome and is defined mathematically as

$$r_A = \frac{1}{2} p_1 + p_2$$

where p_1 and p_2 are probabilities of sharing one and two alleles IBD, respectively. Table 5 also shows also the theoretical IBD distribution and the coefficient of relationship for selected relative pairs.

Table 5. ***Probabilities of sharing zero (p_0), one (p_1) or two alleles (p_2) IBD at autosomal loci and coefficient of relationship r_A for selected relationships which is also the average IBD over the genome.***

Relationship	p_0	p_1	p_2	r_A
Monozygotic twins	0	0	1	1
Dizygotic twins	1/4	1/2	1/4	1/2
Parent-offspring	0	1	0	1/2
Full siblings	1/4	1/2	1/4	1/2
Grandparent-grandchild	1/2	1/2	0	1/4
Half siblings	1/2	1/2	0	1/4
Avuncular	1/2	1/2	0	1/4
First cousins	3/4	1/4	0	1/8

2.3.6 Mapping quantitative trait loci (QTL)

Quantitative trait locus (QTL) mapping refers to the statistical methods that are used to locate and identify loci that influence quantitative traits (QTL). The basic logic underlying QTL mapping is very intuitive, however, the algebra and numerical methods involved are often highly complex and computationally intensive. All methods are based on the underlying principle that family members who have similar trait values should have a higher than expected levels of sharing of genetic material near the QTL (due to linkage) that influence these traits. Conversely, at loci that are not linked to QTL influencing the trait, the degree of genetic sharing should be determined by the degree of relatedness (genetic covariance) and should not correlate with the trait covariance between those individuals (Figure 19).

In human QTL mapping Haseman-Elston regression and variance-components methods are the most popular and are discussed here in detail, but other methods such as score tests and Bayesian methods have also been developed (see Amos[100] and Feingold[101] for reviews). Score tests were initially developed[102,103] to circumvent the normality assumption in Haseman-Elston regression and variance components methods but are less powerful when the trait distribution is close to a normal distribution[104]. The Bayesian methods[105-107] have not been used widely yet because their properties are relatively uninvestigated in human QTL mapping and they are very computationally intensive. However Baesian methods have great advantages over frequentistic methods since they may be used to simultaneously estimate the number and locations of QTL influencing the trait instead of focusing on each putative QTL individually.

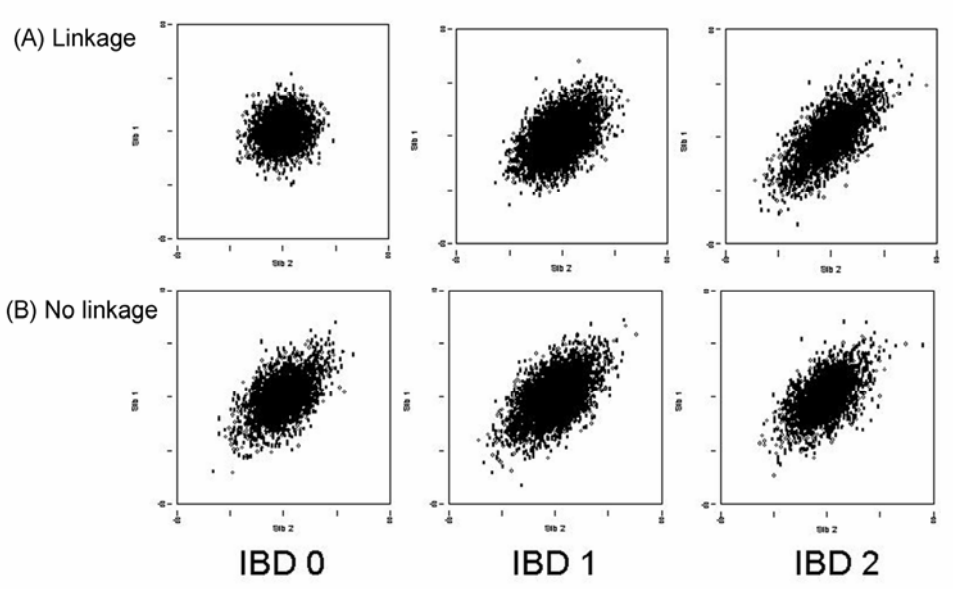

Figure 19. ***Phenotypic covariance and IBD sharing between siblings. In each scatterplot the phenotypic values of the sibs are on the x- and y-axes. (A) The phenotypic covariance (resemblance) is greater in sibs that share more alleles IBD compared to those sibs that share less alleles IBD at this locus. This is the expected scenario if the locus examined is linked to a QTL that influences the trait of interest. (B) There is no marked difference in phenotypic covariance of sibs that share 0, 1 or 2 alleles at the locus examined. Therefore this locus is not considered to be linked to the trait of interest.***

Haseman-Elston Regression

Haseman and Elston were the first to introduce a method for QTL mapping in humans[108]. Their original method was intended for analysis of sib pairs and their approach was based on calculating the squared difference for the sib pair trait values and regressing this upon the estimated proportion of alleles shared IBD (π_i) by the sib pair at the locus being tested (putative QTL). Under the null hypothesis of no linkage, the slope of this regression slope is zero. Under the alternative hypothesis of linkage, the regression slope is negative because similar trait values (that produce small squared differences) should be associated with higher than expected IBD sharing. The original Haseman-Elston regression method is robust in maintaining correct type I error and is computationally simple but has two important limitations 1) it is applicable only to sib pairs and 2) it assumes that the trait is normally distributed in the population. Several extensions have later been proposed to increase statistical power and allow the use of larger families[109-111].

Variance Components Method

The variance components (VC) method is based on modeling the phenotypic covariance between relative pairs as a function of independent additive variance components which are based on the expected genetic covariance and estimated IBD sharing at the locus being tested (putative QTL) and are estimated by maximum likelihood[112-116]. In the basic VC model the phenotypic variance σ_P^2 is modeled as a function of the additive genetic effect σ_q^2 of the marker location (putative QTL), the aggregate additive genetic effect σ_g^2 of all other QTL in the genome (polygenes) and the aggregate environmental effects σ_e^2. Using this model we can test for linkage with a likelihood-ratio test where in the null hypothesis of no linkage σ_g^2 is constrained to zero and in the alternative hypothesis σ_q^2 is a free parameter. Statistical evidence for linkage is then evaluated with a likehood-ratio test which is typically represented as a logarithm-of-odds (LOD) score Z

$$Z = \log_{10} \frac{\max_{\sigma_q^2, \sigma_g^2 \sigma_e^2} L(\sigma_q^2, \sigma_g^2 \sigma_e^2)}{\max_{\sigma_g^2 \sigma_e^2} L(\sigma_q^2 = 0, \sigma_g^2 \sigma_e^2)}$$

where L denotes the likehood of the observed genotype and phenotype data under the alternative ($\sigma_q^2 > 0$) and null hypotheses ($\sigma_q^2 = 0$). The VC framework is applicable to pedigrees of any size and structure and allows for modeling of covariates, household effects, parent-of-origin effects, and multiple QTLs[117,118].

The VC method is more powerful than Haseman-Elston regression[100,101,104] but is computationally very intensive due to extensive use of maximum likelihood methods and is less robust when the trait distribution does not follow a normal distribution producing elevated type I error rates[119]. However, several extensions have been proposed to allow non-normal distibutions in the VC framework that are based on robust estimators[120] and multivariate t-distributions[121] or gamma distributions[122].

2.3.7 Statistical Testing

Gene mapping in non-experimental organisms such as humans is based almost exclusively on statistical inference since controlled experiments are impractical and unethical to conduct - therefore it is necessary for a human geneticist to bear in mind certain fundamental statistical aspects which are discussed in this section. Statistical inference (testing) allows us to objectively evaluate the probability of a given hypothesis; instead of subjectively stating that two samples are different in some aspect we are interested in, statistical testing allows us to assign a level of confidence that the samples are indeed different. All frequentistic statistical testing (in contrast to Bayesian analysis) follows the same basic scheme: 1) formulation of null and alternative hypotheses (H_0 and H_A), 2) calculation of the expected distribution under the null hypothesis, 3) comparing the observed and expected distributions using appropriate test statistics and 4) determining the statistical significance of the difference between the observed and expected distribution by using the known distribution of the test statistic (Figure 20). In the context of genetic mapping usually the hypotheses being tested are:

H_0: "Genetic marker is not linked (or associated) to the trait phenotype."

H_A: "Genetic marker is linked (or associated) to the trait phenotype."

We choose H_0 and H_A in this manner, because we need to be able to calculate the expected distribution under the null hypothesis in a simple manner. If our null hypothesis would be "Genetic marker is linked (or associated) to the trait phenotype." we would require additional parameters to calculate the expected distribution (e.g. recombination fraction, allele frequencies, or magnitude to linkage disequilibrium).

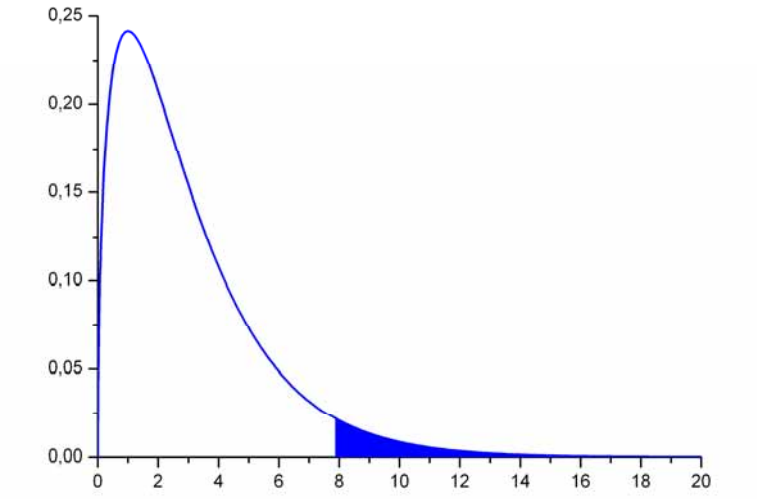

Figure 20. ***The χ2-probability density function with three degrees of freedom. Like all test statistics, the χ2-test statistics follows a known distribution under the null hypothesis and therefore we can evaluate the probability of obtaining the observed test statistic under the null hypothesis. This allows us to evaluate the fit of the observed data to our null hypothesis. For example, with three degrees of freedom, obtaining a χ2 ≥ 7.82 corresponds to p ≤ 0.05 which is equal to the shaded surface area.***

Statistical significance

Statistical significance is a measure of the level of confidence associated to the conclusion of a given statistical test. Statistical significance is usually measured with a *p-value*, which is defined as the probability that chance alone produced the obtained or more extreme test statistic in a given test. Formally p-value is defined as the probability of rejecting the null hypothesis when the null hypothesis is true.

$$p = P(H_0 \text{ rejected} \mid H_0 \text{ true})$$

The p-value is therefore the probability of obtaining a false posititive result i.e. rejecting the null hypothesis inappropriately (type I error, α). There is an inverse relationship between the p-value and the confidence of our conclusion based on the statistical test; small p-values increase our confidence that our conclusion to reject the null hypothesis is correct and not a product of chance (and vice versa).

Statistical power

Statistical power is the probability of obtaining a level of confidence (p-value) that we wish to acquire in our study setting and statistical methods with the sample at hand. Formally statistical power is defined as the probability of rejecting the null hypothesis when the null hypothesis is false (i.e. making the correct conclusion).

$$F = P(H_0 \text{ rejected} \mid H_0 \text{ false})$$

Statistical power relates directly to the chance of a false negative finding (type II error β, note that F = 1-β). Basically we want to assure that we are able to obtain a sufficiently low p-value (e.g. $p < 0.05$) with a sufficiently high probability (e.g. 80%) with our study setting. Determining statistical power can be performed analytically in simple study settings (e.g. case-control studies) and via simulations for complex study settings (e.g. linkage studies with varying family structures).

Multiple Comparisons Problem

Usually the researcher has a subjective pre-set threshold for declaring statistical signicance. In other words this is the probability of an erroneous conclusion (false positive) that the researcher is willing to accept. For example, a researcher may state that he is willing to reject his null hypothesis if the statistical test produces a p-value $p \leq 0.05$. However, we must remember that the p-value is a probability that chance alone produced our test statistic. This translates to the fact that when we use a threshold of $p = 0.05$, one out of every twenty (5%) such tests will produce a false positive result due to chance alone (this is how we define p-value!). Thus, if our experiment involves performing 100 tests, we expect 5 to be declared as significant with a $p = 0.05$ by chance alone. This is known as the problem of *multiple comparisons*. As the number of tests increases, so does the probability that one of the tests will produce a significant result due to chance alone (false positive).

One method of adjusting for multiple comparisons is the Bonferroni-method which is a commonly used method for obtaining the *experiment-wide false positive rate p´*. If our pre-set threshold for significance level is *p* then the probability that a test does not produce a significant result is *1- p*. If we perform *n independent* tests the joint probability that *none* of them will produce a significant results is $(1 - p)^n$. Therefore the probability that *at least one* test will produce a significant result is given by

$$p' = 1 - (1 - p)^n$$

This is the p-value that one should consider significant over the whole range of n tests given a pre-set value of p. If we wish for an experiment-wide false positive rate of p′ (that is, the probability of one, or more, false positives over the entire set of tests is p′), solving for the p-value required for a single test is

$$p = 1 - (1 - p')^{\frac{1}{n}}$$

For the actual obtained p-value, this is the true signicance level that ensures an experiment-wide false positive rate of p' and is known as the Dunn-Sidak method.

In gene mapping studies there are many sources of multiple comparisons such as performing the statistical test for 1) multiple genetic markers, 2) multiple alleles of a genetic marker, 3) multiple phenotypes and/or 4) using multiple genetic models. The proper adjustment for multiple comparisons depends solely on the data set, for example the Dunn-Sidak method is appropriate only if the statistical tests are independent of each other (which is untrue for genetically linked loci). If there is dependence between the statistical tests performed one must use different approaches such as permutation, resampling or simulation techniques for multiple comparison adjustment. The detailed description of the variety of these methods is however beyond the scope of this thesis and can be found elsewhere[123,124].

2.4 Genetics of height

2.4.1 Heritability of height

The earliest heritability estimate for height was 0.57 by Sir Francis Galton which he published in 1889[5] demonstrating a genetic component underlying human height. Since Galton's work, evidence for the influence of genes in height determination has accumulated from hundreds of twin, family and adoption studies.

As noted before, heritability estimates can be obtained by comparing the phenotypic values of relative pairs such as twins or other types of relatives. From twin studies perhaps the most reliable heritability estimates are from analysis of 30.111 Australian and European twin pairs[75] which range between in 0.70-0.94 men and 0.68-93 in women. Large family studies consisting of thousands of families have also estimated the heritability of height is between 0.79-0.98[125,126] in different US populations.Adoption studies are another design that allows the estimation of genetic and environmental influences since siblings (or twins) reared apart share genes but their environment correlate much less compared to siblings reared together.

Also adoption studies show that correlations in height between biological relatives are always higher than with adopted relatives[127-129]. Unfortunately the sample sizes in these studies prevent accurate heritability estimation.

2.4.2 Genome-wide scans for height QTL

Genome-wide screens represent a hypothesis-free approach for localizing genomic regions that influence the trait of interest. Instead of specifying *a priori* some specific candidate genes that may influence the trait we may simply explore the entire genome for statistical evidence for genes influencing the trait. Classical genome-wide screens utilize 400-1000 evenly spaced highly polymorphic genetic markers, typically microsatellites, that serve as genomic landmarks and allow the detection of linkage between the trait phenotype and marker (parametric linkage) or identification of shared genomic segments between relative pairs that resemble each other phenotypically (nonparametric linkage). These linkage screens may implicate positional candidate regions in the genome (i.e. generate hypotheses) which may harbor candidate genes influencing the trait of interest and warrant further investigation by fine-mapping and/or association analysis. Another approach that has become amenable through discovery of enormous amounts of tightly linked SNP markers and inexpensive technologies to genotype them rapidly is genome-wide association analysis (GWA). GWA is also a hypothesis-free design that attempts to directly identify the genetic variants (or variants in very close physical proximity) influencing the trait by utilizing 300,000-500,000 SNP markers across the genome.

To date there have been reports of 17 genome-wide linkage screens and one genome-wide association screen that have focused on adult height (Table 6). As is the story with other multifactorial traits results have been inconsistent and difficult to replicate across different studies. The only genomic regions that have shown overlapping findings (LOD≥2.0) in more than two independent linkage studies are 6q25, 7q35-36, and 9q21-22 while other findings are scattered across the genome.

Only one genome-wide association (GWA) for height has been published to date[130]. The investigators reported convincing association across multiple populations in samples consisting of more than 30,000 unrelated individuals for the HMGA2 gene residing in 12q14. The role of HMGA2 in height determination is also supported by defects in this gene which have been shown to cause gigantism in mice and humans[131,132] and dwarfism[133] in mice. Weedon and colleagues estimated that this gene accounts for 0.3% of height variation in the general population implying that most height genes may be of even smaller effect since the association to HMGA2 was their most significant finding.

Table 6. ***Summary of results of genome-wide linkage screens for height which have reported multipoint LOD scores ≥ 2.0. Studies included in this thesis are not shown here.***

Locus	Peak Marker	deCODE cM (Kosambi)	Multipoint LOD	Population	Number Phenotyped Individuals (Families)	Reference
1p21	D1S1631	125.75	2.25	GENOA African American	611 (na)	Wu et al.. 2003[126]
2q11	D2S113	112.79a	2.23	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
2q24	173 cM (Marshfield)b	173.5c	2.33	Caucasian/Australian	790 (382)	Ellis et al. 2006[135]
3p26	D3S1297-D3S1304	4.94a-19.61	3.17	United Kingdom	1,377 (573)	Wiltshire et al.. 2002[136]
3p14	D3S1766	79.08	2.31	Finland	702 (183)	Hirschhorn et al. 2001[134]
3p21	D3S1300	82.22	3.14	Caucasian/Australian	790 (382)	Ellis et al. 2006[135]
3q23	D3S1764	145.53	2.03	Multiple	6752 (2508)	Wu et al.. 2003[126]
3q26	D3S1763	168.52	2.06	GENOA European American	749 (na)	Wu et al.. 2003[126]
4p16	D4S412	4.42	2.28	European American	1,058 (26)	Liu et al. 2004[137]
4q25	D4S1564	111.18	2.28	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
5p14-p13	D5S2845-D5S1470	41.83a-53.87	2.04	Dutch	477 (174)	Willemsen et al.. 2004[138]
5q31	D5S2115	136.75	2.14	European American	671 (53)	Deng et al.. 2002[139]
5q31	D5S816	138.04	2.26	HyperGen European American	1,100 (na)	Wu et al.. 2003[126]
6q12	D6S1053	81.09	2.66	GENOA European American	749 (na)	Wu et al.. 2003[126]
6q12-q14	D6S1053-D6S1031	81.09a-89.14	2.32	Netherlands	477 (174)	Willemsen et al.. 2004[138]
6q25	D6S2436	161.27	3.06	Netherlands	962 (200)	Xu et al.. 2002[126]
6q25	D6S1007	166.61	3.85	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
6q25	D6S503	185.24	2.45	European American	2,656 (346)	Geller et al. 2003[140]
7p14	D7S484	54.65	2.05	European American	1,816 (79)	Liu et al. 2004[137]
7q21	D7S669-D7S630	89.39a-99.93	2.26	United Kingdom	1,377 (573)	Wiltshire et al.. 2002[136]
7q35	D7S2195	152a	3.4	Sweden	683 (179)	Hirschhorn et al. 2001[134]
7q36	D7S2439-D7S1523	164.44a-164.88a	2.91	Finland	614 (247)	Perola et al.. 2001[141]
7q36	D7S3058	175.87a	2.46	Multiple	6,752 (2,508)	Wu et al.. 2003[126]
8q24	D8S1100-D8S373	148.81-164.47	2.52	Finland	702 (183)	Hirschhorn et al. 2001[134]
9q21	D9S1868	53.04	2.01	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
9p21	D9S301	66.99	2.09	Netherlands	962 (200)	Xu et al.. 2002[126]
9q22	GATA81C04M-	87.58a-101.64	4.34	European American	3,726 (434)	Liu et al. 2006[125]

	ATA18A07M					
9q22	D9S287	98.70	2.74	European American	1,816 (79)	Liu et al. 2004[137]
9q33	D9S1682	128.77	2.22	European American	1,816 (79)	Liu et al. 2004[137]
9q34	D9S1826	157.73	2.66	European American	1,816 (79)	Liu et al. 2004[137]
9q34	D9S1818-D9S1826	150.42-157.73	2.61	Finland	614 (247)	Perola et al.. 2001[141]
11p15	D11S2362-D11S1999	7.27-17.05	2.57	Sweden	683 (179)	Hirschhorn et al. 2001[134]
12p13	D12S341	0.68a	2.07	Finland	702 (183)	Hirschhorn et al. 2001[134]
12q13	D12S1090-D12S398	56.52c-67.42	3.35	Finland	702 (183)	Hirschhorn et al. 2001[134]
13q33	D13S779-D13S797	95.76-105.22	3.56	Finland	702 (183)	Hirschhorn et al. 2001[134]
14q11	GATA74E02A	13.41	2.38	Framingham Heart Study. 31-49 year-olds	792	Beck et al.. 2003[142]
14q11	GATA74E02A	13.41	2.45	Framingham Heart Study. 61-79 year-olds	792	Beck et al.. 2003[142]
14q23	D14S592	60.5	3.67	GENOA European American	749 (na)	Wu et al.. 2003[126]
14q32	D14S1434	100.07	2.01	European American	4,692 (330)	Mukhopadhyay et al. 2003b[143]
15q14	D15S1012	38.12	2.61	African American	580 (221)	Sale et al. 2005[144]
17q21	D17S958	74.32	2.69	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
18q22	D18S1364	91.80a	2.26	European American	4,692 (330)	Mukhopadhyay et al. 2003b[143]
20p12	D20S66-D20S98	38.23-40.76	p = 0.0001	Pima Indians	924/226	Thompson et al.. 1995[145]
20q12	D20S96	66.55	2.51	Finland (Botnia)	379 (58)	Hirschhorn et al. 2001[134]
20q13	81 cM (Marshfield)b	85.0c	2.21	Caucasian/Australian	790 (382)	Ellis et al. 2006[135]
22q12	38 cM (Marshfield)b	42.03c	2.03	Caucasian/Australian	790 (382)	Ellis et al. 2006[135]

Footnotes: a: Genetic distance interpolated using the neighbouring deCODE markers b: Authors did not report peak marker c: Estimated according to Nievergelt et al. 2004 ns: Sequence information not available in the UCSC database Notes: Deng 2002, Liu 2004 and Liu 2006 contain overlapping individuals Mukhopadhyay et al. 2003b also performed analyses modelling imprinting as well as using sex-specific maps, these are excluded from the table Mukhopadhyay et al. 2003a, Mukhopadhyay et al. 2003b, Geller et al. 2003 and Beck et al. 2003 apparently contain overlapping individuals

2.4.3 Gene associations for height

There are hundreds of Mendelian traits which are characterized by growth disturbances (Online Mendelian Inheritance in Man, 2007). However, most of these traits are serious disorders which affect fundamental biological processes and seriously perturb normal development. Thus it is reasonable to assume that the disturbance in growth is a by-product of a severe systemic effect and the genes underlying these traits are unlikely to underlie height variation in the general population. There are however, some Mendelian traits where short stature is a hallmark feature; the defective genes causing these traits may therefore give clues also to the biological mechanisms underlying normal growth and height variation. These traits can be roughly categorized to according to the function of the defective genes to aberrations of the 1) hypothalamic-pituitary-GH-IGF-I axis, 2) sex steroid metabolism and 3) structural components and other factors influencing bone morphogenesis.

Classical growth hormone (GH) deficiency can be caused by mutation of GH releasing hormone receptor (GHRHR), mutations in transcription factors essential for the pituitary GH producing cells (e.g. PIT1 and POU1F1) or mutation or deletion the growth hormone gene (GH1)[146-148]. Missense mutations producing biologically inactive GH, loss-of-funtion of growth hormone receptor (GHR)[149] or defects in the GH signalling cascade (e.g. JAK2, STAT5B or ERK) on the other hand produce GH insensitivity[149-151]. Most of the growth promoting effects of GH are mediated by insulin-like growth factor-1 IGF-1[152] which forms a complex with IGF binding proteins (especially IGFBP3[153]) and the acid labile subunit (ALS)[154] or the IGF-1 receptor gene[155] and thus mutations in these genes may cause GH secondary GH insensitivity (also reviewed by Walenkamp et al.[156] Mullis[148]). To date the only analysis of the association of common variation in genes in the GH-IGF-1 axis and height in the general population[157] reported negative findings and therefore it is still unknown whether variation in these genes contributes to height variation in the general population or not. One gene associated to height that probably acts via the GH-IGF-1 pathway, is one of the receptors (DRD2) of the neurotransmitter dopamine which in turn has multiple roles in regulating appetite and growth hormone secretion. The role of DRD2 was initially discovered from observations of children that are chronically exposed pre- and postnatally to dopamine receptor D2-blockers who tend to increase in height[158]. Variations in DRD2 have also been associated with height in the general Japanese population[159].

Sex steroids play a pivotal role especially in pubertal growth and although single gene mutations are quite rare there are two noteworthy examples of defects in sex steroid metabolism in respect to growth. Mutations in the estradiol receptor ESR1 causing estrogen resistance lead to tall stature because this reduces the response of bones to estradiol which is the key accelerator of bone maturation and epiphyseal closure. Due to this lowered estradiol response the patients continue growing until their late twenties or even later[160] resulting in increased stature. Another gene implicated in abnormal stature functions via sex steroid pathways is CYP19 which codes for aromatase that is responsible for converting androgens to estrogens. Certain loss-of-function mutations in aromatase result in estrogen deficiency and tall stature especially in males[161,162]. On the other hand, gain-of-function mutations in CYP19 causes excess estrogen which in boys causes gynecomastia, a premature growth spurt, early fusion of epiphyses, and decreased adult height[163]. Polymorphisms in CYP19 have also been associated to variation in height in the general population[41].

Bone morphogenesis is a highly complex process where there are tens of known genes that when defected cause skeletal dysplasias and dysosotoses (reviewed extensively by Superti-Furga et al. [164] and Kant et al.[165]). Interestingly over half of all known mutations underlying abnormal bone formation are located within FGFR3 and COL2A1[165]. These skeletal dysplasias and dysostoses produce dramatic phenotypes and currently it is unknown how they relate to height in the general population. To date, only one gene involving bone metabolism, the vitamin D receptor has been associated to height in the general population. The vitamin D receptor is a nuclear hormone receptor which mediates the action of vitamin D3 (calcitriol) by inducing the synthesis of osteocalcin, the most abundant noncollagenous protein in bone. Polymorphisms in VDR have been shown to associate with adult height in Japanese and European-Americans[166-168] and many bone phenotypes across multiple populations[169-172].

One of the most convincing single gene findings underlying idiopathic short stature (short stature of unknown cause) that does not fit into the rough categories stated above are mutations within the SHOX gene located in the pseudoautosomal region of the sex chromosomes[173,174]. SHOX encodes for a transcription factor belonging to the homeobox gene family of developmental regulators. SHOX haploinsufficiency has been suggested as the cause of the short stature phenotype in Turner syndrome patients[44].

Table 7. ***Selected genes that have influence on height***

Gene name	Gene Symbol	Locus	Function
Growth hormone receptor	GHR	5p12	Mutations in GHR have been associated with Laron syndrome, also known as the growth hormone insensitivity syndrome (GHIS), a disorder characterized by short stature.
Estrogen receptor 1	ESR1	6q25	Nuclear hormone receptor. Loss-of-function of ESR1 leads to estrogen resistance and tall stature due to incomplete epiphyseal closure.
Growth hormone releasing hormone receptor	GHRHR	7p15	Mutations in this gene have been associated with isolated growth hormone deficiency (IGHD), also known as Dwarfism of Sindh, which is characterized by short stature.
Insulin-like growth factor binding protein 3	IGFBP3	7p13	Mutations may cause biologically inactive IGF-1 resulting in short stature.
Janus kinase 2	JAK2	9p24	Involved in growth hormone signalling.
Dopamine D2 receptor gene	DRD2	11q23	Receptor of dopamine which plays a major role in the regulation of appetite and growth hormone.
Insulin-like growth factor	IGF1	12q23	Mediates many of the growth-promoting effects of growth hormone. IGF1 deficiency is an autosomal recessive disorder characterized by growth retardation, sensorineural deafness and mental retardation
Vitamin D3 receptor	VDR	12q13	Nuclear hormone receptor that mediates the action of vitamin D3 by controlling the expression of hormone sensitive genes
Cytochrome P450, family 19	CYP19A1	15q21	Involved in estrogen biosynthesis. CYP19A1 mutations have been associated with height via estrogen dosage.
Insulin-like growth factor 1 receptor	IGF1R	15q26	Loss-of-function mutations lead to IGF-1 insensitivity and pre- and postnatal growth retardation.
Acid-labile subunit	IGFALS	16p13	Mutations may cause biologically inactive IGF-1 resulting in short stature.
Signal transducer and activator of transcription 5B	STAT5b	17q21	Involved in growth hormone signalling. Defects in STAT5B are the cause of Laron type dwarfism II (LTD2) which is characterized by growth hormone insensitivity.
Short stature homeobox-containing gene	SHOX	Xp22/Yp11	Involved in idiopathic growth retardation and in the short stature phenotype of Turner syndrome patients

.

3 AIMS OF THE STUDY

The aim of this study was to localize and identify genetic variants that influence the determination of adult height in the general population by addressing the following specific aims:

1) To develop methods and protocols for the efficient use of samples originally ascertained for different traits by various investigators in combined genome-wide linkage analysis (I, II, III).

2) To perform combined data analyses of genome-wide linkage screens performed in Finnish family samples and investigate the potential of sex-limited analysis in order to locate sex-specific genetic effects (I).

3) To perform combined data analysis of genome-wide linkage screens performed in Australian, Danish, Finnish, Dutch, Swedish and United Kingdom families and investigate also the usefulness of reducing environmental variation in genome-wide analyses by restricting analyses to dizygotic twin pairs (II).

4) To carry out combined data analysis of genome-wide linkage screens performed in families from the United States and investigate the role of within country population variation by performing joint and individual analyses on families of African-American and European-American origin across the cohorts (III).

5) To study the potential sex-specific genetic architecture of height (I, II, III).

4 MATERIALS AND METHODS

4.1 Study subjects

The reader is referred to the original publications (I-III) and the references therein for a more detailed description of the studies.

4.1.1 The Finnish family sample

In Study I all genome-screened families originated from Finland and they were ascertained for type 2 diabetes, familial low-HDL cholesterol or familial combined hyperlipidemia (FCHL). The total study sample contained 1,417 individuals from 277 families with genome-wide genotype data and phenotype information. Details of the sample used in Study I are shown in Table 8.

Table 8. ***Demographic characteristics of the sample used in Study I.***

		Stature (cm)			Age
	n	Mean	SD	Range	Mean
Botnia					
Males	211	174.7	5.9	157.0-193.0	58.7
Females	205	161.2	5.9	143.0-181.0	58.2
Helsinki					
Males	288	174.4	6.7	153.7-197.0	56.6
Females	380	160	6.3	136.0-184.0	60.1
FCHL					
Males	63	175.7	5.9	164.0-191.0	47
Females	76	161.5	6.2	147.0-176.0	51.4
Low-HDL					
Males	97	176.7	6.2	162.0-190.0	49.3
Females	97	164.2	6.2	149.5-170.0	50
Combined					
Males	659	174.7	6.4	153.7-197.0	55.4
Females	758	161	6.3	136.0-184.0	57.6

4.1.2 Families from Australia and Europe

The family samples analyzed in study II were derived from twin cohorts provided by the GenomEUtwin consortium (http://www.genomeutwin.org) which consists of eight twin cohorts (Australian, Danish, Dutch, Finnish, Italian, Norwegian, Swedish, and the United Kingdom) with the total resource of hundreds of thousands of twin pairs. The total number of individuals with genome-wide genotype data and phenotype information was 8,450 individuals from 3,817 families.

Details of the sample used in Study II after diagnostic analyses of phenotype data were performed are shown in Table 9.

Table 9. ***Demographic characteristics of the sample used in Study II.***

		Stature (cm)			Age
	n	Mean	SD	Range	Mean
Australia					
Males	1005	177.2	6.5	152-203	40.9
Females	1604	163.4	6.6	144-186	42.2
Denmark					
Males	248	176.8	7.9	152-199	53.7
Females	380	164.1	7.0	144-183	61.1
Finland					
Males	509	175.6	6.1	157.5-194	52.6
Females	342	160.8	6.4	147-176	61.2
Netherlands					
Males	481	179.3	6.7	162-203	40.2
Females	605	168.7	6.3	147-189	37.9
Sweden					
Males	531	175.4	5.9	156-193	74.6
Females	533	162.6	6.0	137-163	75.0
United Kingdom					
Males	-	-	-	-	-
Females	2212	162.5	6.4	141-191	47.3
Combined					
Males	2774	177.5	6.8	152-203	50.2
Females	5676	163.4	6.6	137-191	49.1

4.1.3 Families from the United States

In Study III we analyzed African-American and European-American families ascertained for type 2 diabetes, blood pressure-related traits, obstructive sleep apnea and systemic lupus erythematosus. The combined sample consisted of 9,306 individuals from 3,032 families with genome-wide genotype and phenotype information. Descriptive statistics of this sample are shown in Table 10.

Table 10. ***Demographic characteristics of the sample used in Study III.***

		Stature (cm)			Age
	n	Mean	SD	Range	Mean
African-American					
Males	1617	177	6.6	158-196	50.4
Females	3161	163.6	6.3	144.8-182	51.4
European-American					
Males	2183	176.4	6.5	158-195.6	54.1
Females	2884	162.4	6.2	144-180.3	53.7
Combined					
Males	3550	176.7	6.6	158-196	52.5
Females	5756	163.1	6.3	144-182	52.5

4.2 Methods and statistical analyses

4.2.1 Combining the genome-wide screens

Marker map construction

The advent of commercial multi-allelic marker sets (Weber sets, ABI sets) in the mid-1990s was essential for efficient genome-wide linkage screening. However, each genome-wide screen is essentially performed using a unique set of genetic markers (marker map) because for example some markers in a commercially available marker set may not be sufficiently polymorphic in some study samples, or genotyping success rate is low. Also many investigators follow-up on interesting genomic regions with additional genetic markers or add markers simply to ensure sufficient coverage of a given region.

Building a common genetic marker map is essential for combining data across multiple genome-wide screens so that marker genotypes can be assigned to their correct positions. Also, since the correct genetic map is crucial for power and accuracy of multipoint linkage analysis[175-177] we applied the most current data and strict criteria in genetic map building.

Combining marker maps however, is not trivial due to many factors. First there is no standard nomenclature for multi-allelic markers and they may have up to ten aliases; for example the marker ATA29D04 present in one screen may be named D1S1631 or RH27948 in other screens. Second, published reference genetic maps such as the Marshfield[178] and the deCODE map[85] disagree about some locations; for example they place markers D1S2783 and D1S195 in different orders along the chromosome. Third, genetic maps also differ in map resolution (number of meioses used to build the map), marker density and choice of markers.

We used the genome sequence (physical map) to assign the marker order, because it is based on physical experiments rather than statistical inference (which is the basis of marker positioning in genetic maps). However, since recombination frequency is not constant over the genome we also needed to rely on a reference genetic map to obtain genetic locations. We chose the deCODE map because out of the published genetic maps it is statistically most reliable and has the highest resolution because it is based on the largest number of informative meioses (1,257 versus 188 in the Marshfield map).

Our strategy for building the common genetic map is illustrated in Figure 21 and is as follows: 1) obtain a unique reference name and physical location (bp) for each marker from the University of California Santa Cruz (UCSC) database (http://genome.ucsc.edu) and order the markers based on their physical locations, 2) obtain the genetic location (cM) for each marker from the deCODE genetic map and 3) for those markers which were not included in the deCODE marker map or were in disagreement with the sequence information, use linear interpolation for genetic location estimation using the physical and genetic locations of the immediately neighboring deCODE markers. For this purpose we developed publicly available software Cartographer (https://apps.bioinfo.helsinki.fi/software/cartographer.aspx).

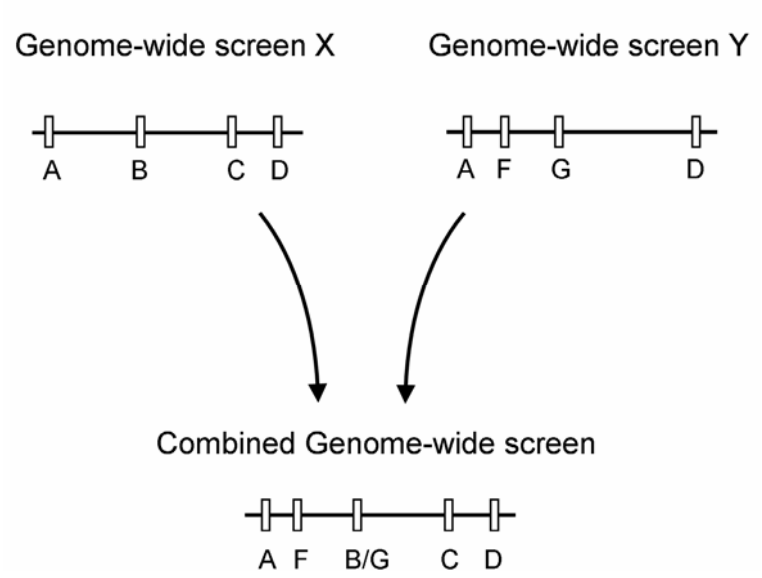

Figure 21. ***Schematic representation of combining marker maps across multiple genome-wide screens. The genetic map of genome-wide screen X includes genetic markers A, B, C, and D while genetic map of genome-screen Y includes markers A, F, G, and D. Querying the UCSC database identifies markers B and G as the same marker, although it is genotyped under different names in the two genetic maps. The physical locations of the markers are also retrieved from the UCSC database and the markers are ordered accordingly. Last, the genetic locations are retrieved from the deCODE genetic map and interpolated using their physical locations if the markers are not included in the deCODE map.***

Combining genotype data

Once a common marker map is established combining the genotype data is a trivial yet tedious task of assigning genotypes to proper locations along the map. For this purpose we wrote custom software Mergescan (Sammalisto and Hiekkalinna, unpublished). However, because the genome-wide screens were performed at different genotyping centers and different batches we could not treat markers from different genome-wide screens as identical even if they have been genotyped under the same name since they may differ in allele coding. It is common practice to code alleles 1, 2, 3… n based on their frequencies in the sample; therefore allele 1 for a given marker in study A may correspond to allele 2 for that marker in study B. One could try to resolve this by recoding alleles in each sample according to their frequency before combining; this however would assume that the proportions of allele frequencies are the same in each study sample which is clearly an unsafe assumption. The allele coding problem can be resolved either by 1) assigning unique allele numbers for a marker in each sample or by 2) treating markers from different samples as different markers by setting some small arbitrary genetic distance between them (e.g. 0.0001 cM). The former is more suitable for two-point analysis while the latter is more suitable for multipoint analysis. Because in this study we were primarily focused on the multipoint analysis we chose the latter strategy.

4.2.2 Statistical analysis

The methods described here were used throughout the study (I-III), therefore there are no indications of which methods were used in which part of the study.

Familial relationship verification

The genotype quality control methods and linkage analysis methods assume that family relationships are correctly specified. In some cases such as non-paternity, also violations to Mendelian transmission of alleles at multiple loci may uncover incorrectly specified familial relationships. However, for more distant relationships simply focusing on Mendelian transmission is unlikely to uncover all relationship errors. In the context of genetic mapping knowledge on the true relationships is essential since all genetic analyses are performed conditional on the family structure. For example, the expected genetic covariances utilized in the variance components linkage analysis are derived from the declared familial relationships.

In this study Graphical Representation of Relationships, GRR[179] was used to perform the familial relationship testing. GRR implements a simple general method that plots the mean and standard deviation of genome-wide IBS sharing between all pairs of individuals in a given sample resulting in distinct clustering of specific types

of relatives and unrelated pairs. The approach implemented in GRR has two clear advantages over many other approaches 1) it does not require specification of parameters (such as allele frequencies) and 2) since no inferential statistics are used there is no need to set significance levels or be wary of multiple comparisons. GRR is designed for genome-wide studies using multi-allelic markers and it can be considered as accurate when individuals are genotyped for at least 50 loci. Figure 22 GRR shows how potential errors can be detected from the GRR output. If suspicious relationships were detected the original investigators of the corresponding genome-wide screen were contacted. If they were unable to provide information that would allow us to resolve the suspected error either the pair in question or the whole family was discarded prior to genetic analyses.

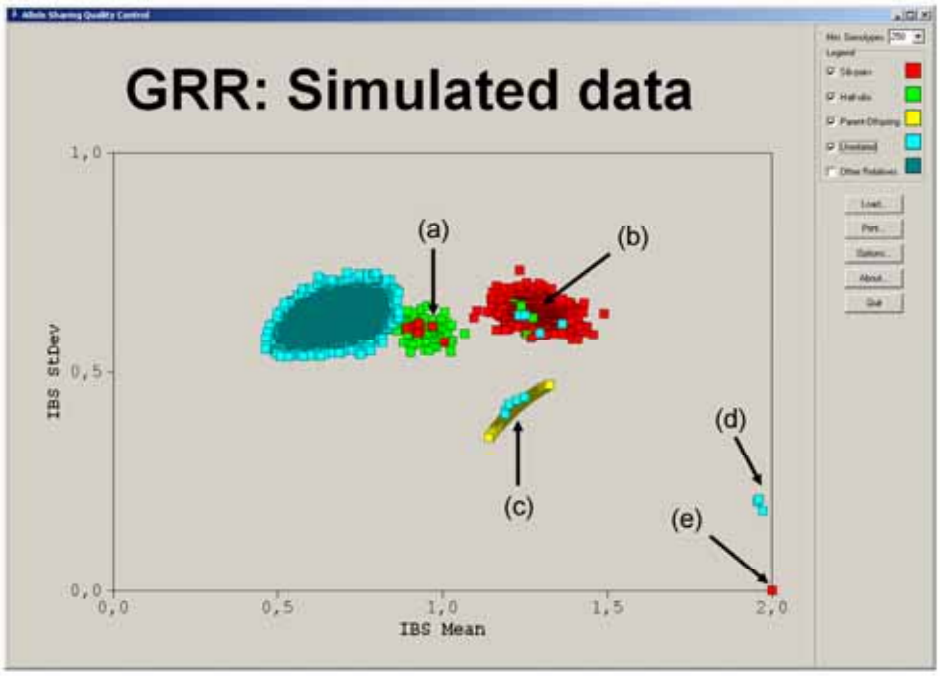

Figure 22. ***GRR output for a simulated data set. Different declared relationships, sib-pairs, half-sibs, parent-offspring and unrelated pairs are shown in different colours and form distinct clusters when relationships are correct. The results show many types of likely relationship errors: a) sibs who are really half-sibs, b) half-sibs and unrelateds which are really sibs, c) unrelateds which are parent-offspring pairs, d) unrelateds that may be either genotyping duplicates or MZ twins and e) a sib-pair which is really a MZ pair. Simulated data is provided by GRR authors (www.sph.umich.edu/csg/abecasis/GRR/)***

Genotype quality controls

Since all the genotypes were already produced prior to this study, genotype quality could only be assessed by statistical methods. Elimination of all Mendelian inconsistencies is essential prior to genetic analyses. In this study we applied Pecheck 1.1[180] for singlepoint analysis of Mendelian inconsistencies in the data. If inconsistencies were detected we removed the genotypes for that marker in all family members using the automatic routines (option "-z") implemented in Pedcheck. Figure 23 shows an example of a Mendelian inconsistency due to a common type of genotyping error called "allele dropout".

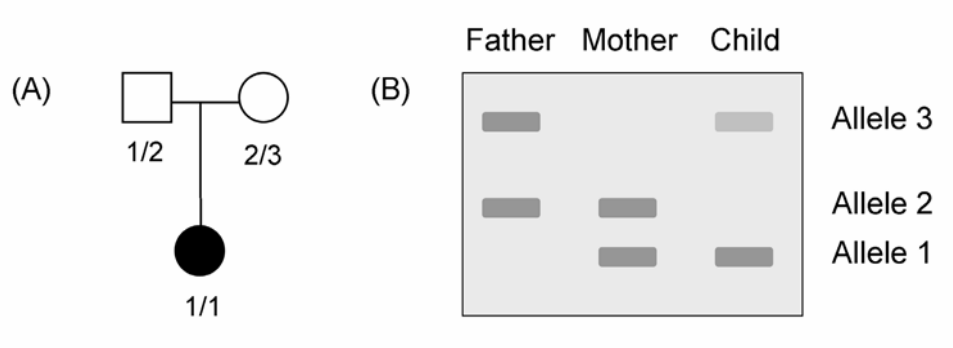

Figure 23. ***(A) A family with father, mother and a child who are genotyped for a genetic marker. The genotypes however, are Mendelian inconsistent since the homozygous child must have inherited allele 1 from both parents but the mother does not carry any copies of allele 1. (B) Electrophoresis diagram from the genotyping procedure used to genotype this family. The true genotype of the child is 1/3 but for some reason allele 3 is below the detection threshold and thus the child is erroneously declared as 1/1 homozygote. This phenomenon is called allele dropout (the presence of allele 3 is not detected) and is a common type of genotyping error.***

Multipoint linkage analysis is highly sensitive to genotyping error and it has been shown that error rates as low as 1% can decrease the power to detect loci significantly[181,182]. Singlepoint analysis may not be able to uncover all inconsistencies in the genotype data for example due to missing parental genotypes or uninformative markers (e.g. single SNPs). Multipoint genotyping error detection methods examine multiple linked markers simultaneously and are able to detect genotyping errors by detecting excessive recombination (Figure 24). Multipoint error detection has been shown to greatly improve the power of linkage analysis[182] and it is especially relevant for small families where the majority of genotyping errors are Mendelian consistent[183,184] i.e. not detectable using singlepoint methods.

We applied the multipoint method for genotyping error detection implemented in Merlin 1.0.1[185] and removed all genotypes flagged by Merlin as unlikely using the Pedwipe utility program distributed with the Merlin package.

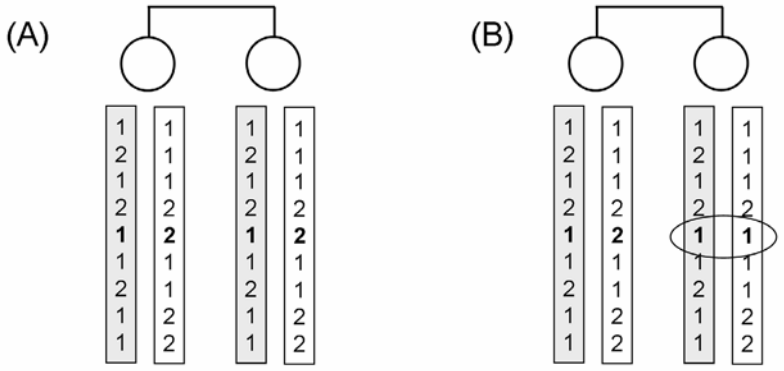

Figure 24. ***Concept of multipoint error detection. Sibs that are genotyped for nine closely linked bi-allelic genetic markers. (A) It is likely that sibs have inherited identical haplotypes from each parent since they have identical genotypes at all loci. (B) If the genotype at one of the marker is different (circled) the genotype at this locus contradicts the inheritance pattern from all other linked markers since it would imply one of the parents carried two nearly identical copies of the chromosome or two recombination events occurred. MERLIN can compute the likelihood of each scenario and evaluate their likelihoods allowing the detection of genotyping errors. Figure reproduced from the Merlin web tutorial (www.sph.umich.edu/csg/abecasis/Merlin)***

Most genetic analysis methods assume that genotypes at genetic markers are in Hardy-Weinberg equilibrium (HWE) and therefore the validity of this assumption should also be verified. Since it has been shown empirically that HWE holds for most human populations, deviations from the expected HWE proportions may suggest problems with genotyping or undetected population structure (stratification) or in samples ascertained for manifestation of a certain disease, an association between the marker and disease susceptibility. In linkage analysis the HWE assumption is used to assign the probability of compatible founder genotypes when they are missing. In this study we used Pedstats[186] to test for HWE equilibrium using the "checkunrelated"-sampling scheme that performs the test by sampling individuals from the families whose genotypes are independent of each other. If a marker showed deviation from HWE at a statistically significant level (experiment-wide $p \leq 0.05$) we ignored this marker in subsequent genetic analyses.

Statistical diagnostics of the phenotype data

Parametric statistical methods such as Pearson's correlation, linear regression and analysis of variance (ANOVA) assume certain characteristics of the variables analyzed and may produce highly biased estimates, false positives and/or false negatives if these assumptions are violated (see Figures 25 and 26 for examples). Since the variance components method used for the linkage analyses in this study is essentially an extension of multiple regression analysis it assumes that the trait analyzed follows a normal distribution, therefore the validity of this assumption and the presence of outliers (extreme trait values) must be checked with statistical diagnostics methods. It has been shown that violation of the normality assumption and/or relatively few outliers even in large data sets may elevate the false positive rate of the variance components linkage analyses substantially[119,187,188]. We performed basic diagnostics using SPSS 11.0 to ensure that height was normally distributed in our sample and excluded outliers that deviated more than three standard deviations from the sex-specific mean in that sample. Also, the phenotypes of all individuals less than 23 years old were excluded because these individuals may still be growing.

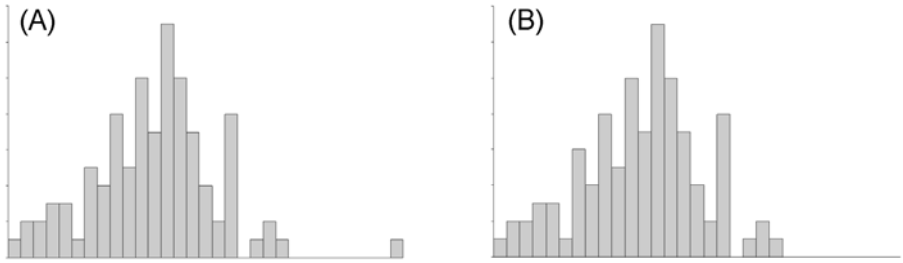

Figure 25. ***Effect of outliers on distribution measures. Histograms A and B are based on 100 data points but in A there is a single outlier in the data. Removing this outlier reduces the mean from 0.10 to 0.05 (50% reduction) and the variance from 0.95 to 0.80 (19% reduction) even though only 1% of the data is altered.***

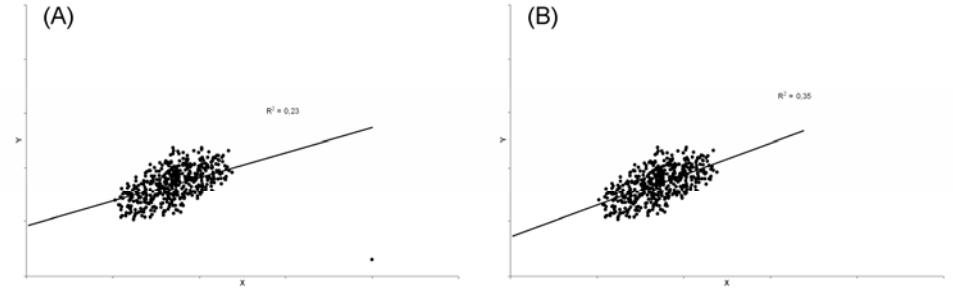

Figure 26. ***The effect of outliers to regression. Scatterplots A and B are based on 500 data points but in A there is a single outlier in the data. Removing this outlier increases the regression coefficient from 0.23 to 0.35 (52% increase) even though only 0.2% of the data has changed.***

Height measurement

For most samples body height was measured at health care centers without shoes, however for some samples only self-reported height was available. Although there is upward bias in self-reported height it has been shown in Finnish twins that the correlation between measured and self-reported height is very strong 0.98 for men and 0.96 for women[62]. Secondly since this bias is systematic, i.e. virtually everyone overestimates their height, it has little effect on height variance which is the primary focus of interest in variance components linkage analyses. In some of the US genome-wide screens height was measured in inches which were converted to centimeters using 1 inch = 2.54 cm.

Variance components linkage analysis

For the the linkage analyses we used the variance components method implemented in Merlin 1.0.1.[185]. Merlin uses the Lander-Green algorithm[189] for the IBD estimation and standard variance components framework for heritability analysis and linkage analysis (see 2.6.3 for mathematical description) assuming no dominance or epistatic variance. Due to its use of the Lander-Green algorithm Merlin is designed to provide exact likelihood calculations for small families (≈20) that are genotyped for large numbers of linked markers. Merlin utilizes fast algorithms based on sparse gene flow trees for IBD estimation which improves analysis efficiency considerably in terms of computation speed and memory consumption compared to similar programs such as Genehunter[190] or Allegro[191].

The VC method implemented in Merlin allows the incorporation of covariates in the variance components model. Covariate adjustment is used to improve the power to detect QTL by reducing the phenotypic variance that is not attributable to genetic effects[192]. If the covariate does not correlate with the QTL influencing the trait of interest, covariate adjustment of the trait value increases the power to detect that QTL. Conversely, if the covariate and the QTL do correlate (or in the extreme case there is a causal relationship), then the power to detect the QTL is decreased[192]. If we wish to find genes for height, we want to eliminate all sources of variation in height that are not due to genes. For example, we know that sex is an important source of "environmental" variation in height so when we want to analyze height in a sample consisting of males and females we eliminate the effect of sex by performing sex-adjustment. Basically covariate adjustment in the VC framework is performed by using multiple regression analysis to eliminate the effect of that covariate to the phenotype of interest.

In our analyses we used sex, age and cohort as covariates in the VC model. Since height is a sexually dimorphic trait (the mean heights of males and females is significantly different), sex is an obvious covariate. Furthermore as the age range of our samples is wide we also incorporated age as a covariate to control for the secular trend. Because there were slight differences in the distribution of height between samples we combined we also included a cohort identifier as a covariate in our analyses. This was most evident in Study II where we combined several distinct populations (Table 9).

It has been shown that inclusion of ungenotyped and unphenotyped individuals that are not required for specification of pedigree structure introduces noise in the variance-components analyses[193] because the IBD estimates for ungenotyped individuals will be based on *theoretical* expected genetic covariance (see the section "Genetic covariance between relatives") instead of genotype data and may therefore be inaccurate. For this reason, we used Merlin's "trim" option to automatically remove these individuals prior to genetic analysis. Also, all variance components analyses were automated using AUTOGSCAN[194].

Stratification strategies

Sex can be considered as an environmental factor that can modify the penetrance or expressivity of a trait[8,195,196]. In addition to using sex as a covariate to correct for the difference in average height between males and females we also stratified our sample according to sex (i.e. analyzed males and females separately). This was done because as noted before, covariate adjustment reduces the power to detect linkage if the covariate and the QTL examined correlate[192]. Therefore if there are QTL that are influenced by sex or are sex-specific there may be more power to detect them in sex-specific analyses. These sex-specific analyses were performed by ignoring the phenotype data of one sex while retaining all available genotypes in order to maximize IBD information in the sample.

In study II where all families were ascertained on the basis of monozygotic or dizygotic twin pairs we also aimed to reduce within-pair environmental variation by only including the phenotypes of the dizygotic twin pairs. Utilizing dizygotic twins instead of siblings reduces environmental variation due to many factors: 1) they are of the same age, 2) they share fetal conditions and 3) they are likely to share childhood environment more closely than siblings of different ages[197]. These analyses were also performed by retaining all available genotypes for maximum IBD information.

We also performed linkage analyses for each study sample separately in each study I, II and III for two reasons: 1) combining samples may increase genetic and environmental heterogeneity and thus reduce the power to detect QTL and 2) comparison of the subset and the combined analyses allowed us to evaluate the gain in statitistical power due to combined analysis of the subsets.

Empirical significance determination

The significance of an obtained test statistic is assessed by comparison to the known theoretical distribution of that type of test statistic. However, in the case of LOD score analysis the distribution under the null hypothesis (no linkage) depends on many factors (marker density and informativeness, family structure, patterns of missing data etc.) so the theoretical distribution may not hold for all genome-wide screens. Therefore in order to assess the findings of a specific genome-wide screen ideally one should derive the empirical distribution of the test statistic in that specific screen via simulation. In this study we used the gene-dropping simulation method implemented in Merlin 1.0.1 to randomly generate genotypes (under the null hypothesis of no linkage) conditional on the genetic map and allele frequencies while retaining the family structures, phenotypes and covariates of the individuals as well as missing genotype data patterns.

The gene-dropping approach is considered as the gold standard for determining genome-wide significance, but due to its heavy computational burden many alternative approaches have also been developed[198-200]. Since these simulated data sets are comparable to the original data in all other aspects except that any linkage is due to chance, we are able to evaluate the probability that a given finding in the original data is due to chance alone (false positive). The logic of evaluating the empirical significance is simple and is shown in Figure 27: 1) analyze the original (real) data, 2) record the reference (e.g. highest) LOD-score obtained for the original data, 3) use gene-dropping to simulate a replicate data set, 4) analyze the replicate data set exactly like the original data, 5) examine if reference LOD-score or higher was observed in the replicate, 6) repeat steps 2-4 e.g. 100 replicates or more, depending on computational resources and 7) determine the empirical significance of the reference LOD score by dividing the number of replicates where the reference LOD-score or higher was observed by the total number of replicates. For example, if one simulates 100 replicate data sets and in 4 of them the reference LOD score (e.g. LOD=3.00) is reached or exceeded, then the empirical p-value for the refrerence LOD score is p = 4/100 = 0.04.

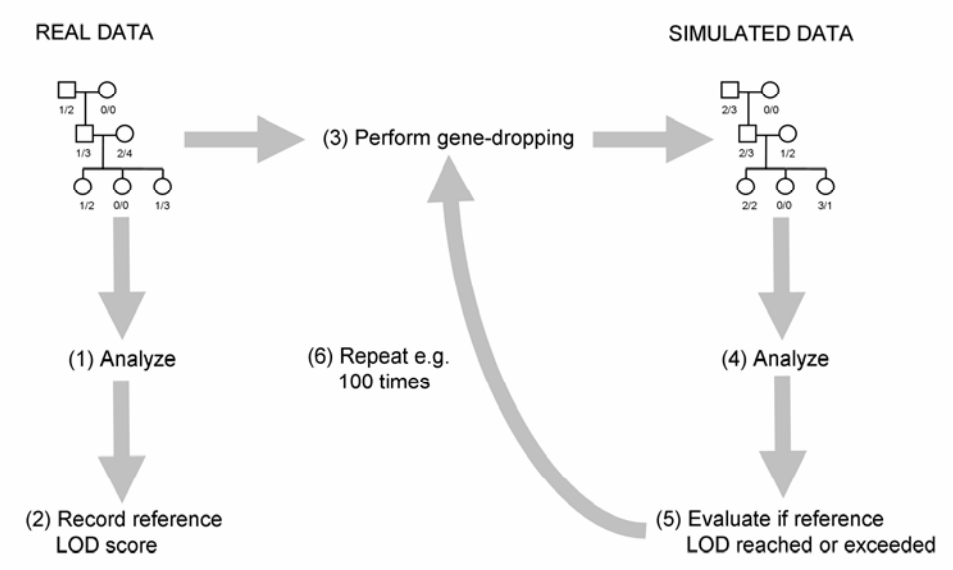

Figure 27. ***Schematic representation of determining empirical signicance of a LOD score by gene-dropping simulations. Note that the retainment of the missing data patterns: in the real data the mother in the top generation and one of the sibs in the bottom generation are not genotyped therefore these missing data patters are retained in all replicates. Calculation of the empirical significance is not shown in this diagram.***

5 RESULTS

5.1 Combined genome-wide screen in Finnish families

5.1.1 Stratification of families

The Finnish families used in this study were originally ascertained for type 2 diabetes, familial combined hyperlipidemia or familial low-HDL levels and in total comprised of 1,417 individuals from 277 families with genome-wide genotype and phenotype data. Most families were extended pedigrees ranging from two to five generations (average 2.8) and consisted between 2 to 15 informative members (average 4.9).

5.1.2 Heritability estimates

The covariate-adjusted (age, sex and cohort) narrow sense heritability of stature was 0.84 in the combined sample. In the sex-stratified analyses age and cohort were used as covariates and the heritabily estimates for stature were 0.95 for males and 0.99 for females.

5.1.3 Genome-wide linkage results

Genome-wide multipoint linkage results are shown in Figure 28. Six chromosomal regions were linked to stature with multipoint LOD scores ≥ 2.0 (Table 11). Only the male-specific locus on 1p21 was statistically significantly linked to stature according to 100 simulated genome scans with an empirical p-value of $p < 0.01$ (Wilson's 95% confidence interval[201] 0-0.037).

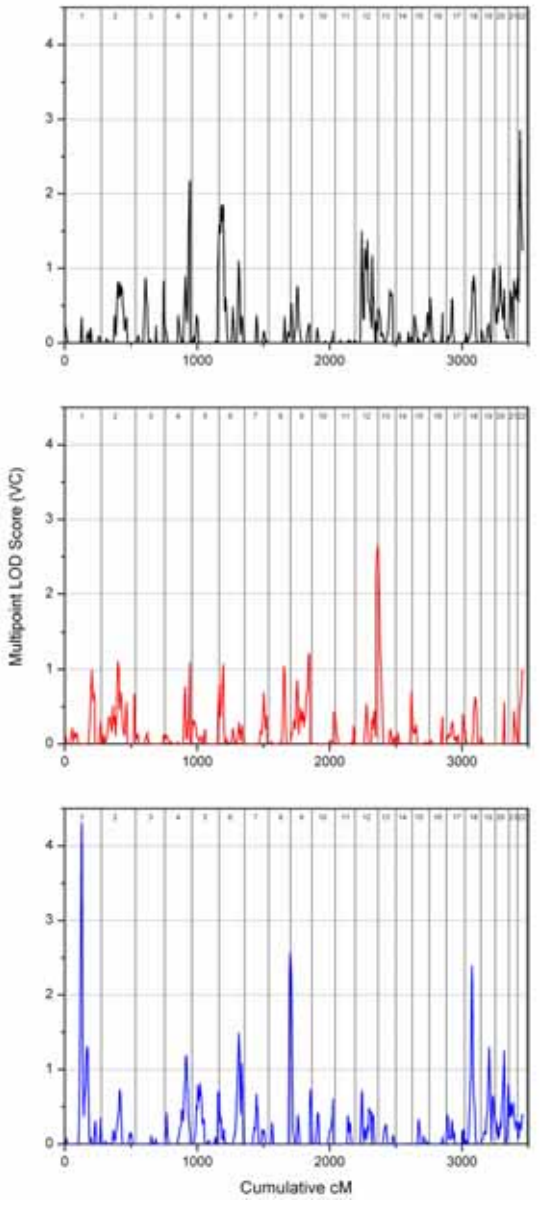

Figure 28. ***Multipoint LOD scores for the Finnish families. The results for the joint analysis of males and females, females and males are shown from top to bottom, respectively.***

Table 11. ***Multipoint LOD scores ≥ 2.0 observed in the Finnish families. Asterisks denote interpolated genetic distances.***

			Multipoint LOD Score		
Locus	Peak Marker	deCODE cM	Males	Females	Combined
1p21	D1S1631	125.75	**4.25**	0.00	0.34
4q35	D4S426	202.69	0.17	1.06	**2.18**
9p24	D9S2169	12.55*	**2.57**	0.00	0.15
13q12	D13S221	16.26	0.00	**2.66**	0.47
18q21	D18S60	86.76*	**2.39**	0.00	0.57
22q13	D22S282	50.81	0.23	0.47	**2.85**

5.2 Combined genome-wide screen in Australian and European twin families

5.2.1 Stratification of families

The Australian, Danish, Finnish, Dutch, and UK families derived from the twin cohorts of the GenomEUtwin consortium provided genome-wide genotype and phenotype data also for non-twin family members in addition to the twin pair (MZ or DZ) for which they were ascertained on. Since sample size is a critical determinant of statistical power in genetic mapping studies we performed the variance components analysis also using all available data even though the initial premise of the study design was minimizing envirionmental variation by utilizing DZ twins. From the available 3,817 families we had data for 8,450 individuals in the former and 6,602 in the latter sample.

5.2.2 Heritability estimates

When all available data was analyzed, the covariate-adjusted (age, sex and cohort) heritability estimates for stature were 0.82 in the combined analyses, 0.98 in the females- and 0.93 for males-only analyses. When only the DZ twin pairs were analyzed, the heritability estimates were 0.99 in the combined analyses, 1.0 in the females- and 1.0 for males-only analyses. Only age and cohort identifier were used as covariates in the sex-stratified analyses.

5.2.3 Genome-wide linkage results

Genome-wide multipoint linkage results are shown in Figure 29. In this sample six chromosomal regions showed evidence for linkage to stature, with LOD scores ≥ 2.0 (Table 12). However, none of the loci retained empirical significance according to

our simulation; the empirical p-value for the strongest linkage on 8p21 (multipoint LOD = 3.28) was p = 0.08 (Wilson's 95% CI 0.04-0.15).

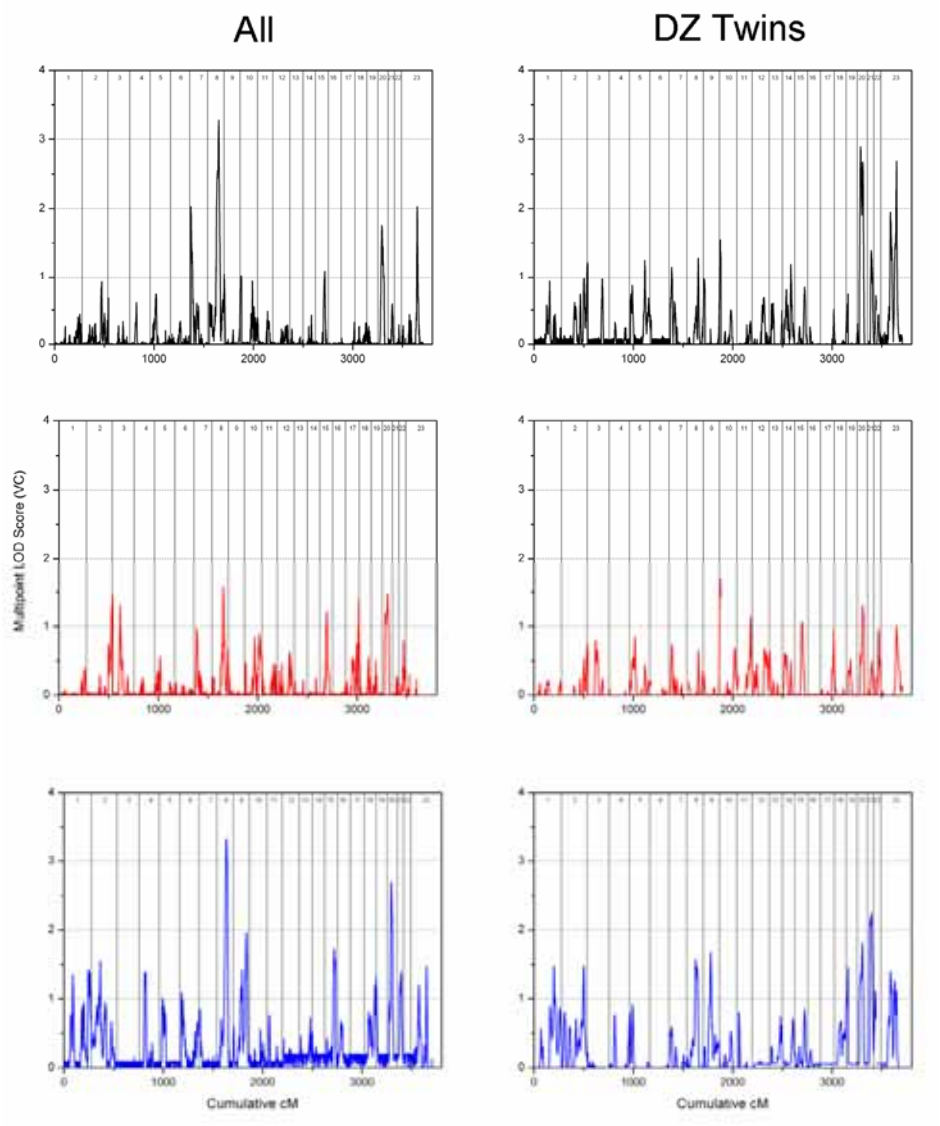

Figure 29. ***Multipoint LOD scores for Australian and European families. The left panel displays the results using all available individuals and the right for DZ twins only. The results for males and females jointly as well as, females and males separately are shown from top to bottom, respectively.***

Table 12. ***Multipoint LOD scores ≥ 2.0 observed in the Australian and European families***

		Multipoint LOD Score					
		All individuals			DZ twins only		
Locus	deCODE cM	Males	Females	Combined	Males	Females	Combined
7p22	1	0.78	0.00	**2.03**	0.52	0.05	0.00
8q21	97	1.07	1.58	**3.28**	0.51	0.57	1.25
20p13	15	1.70	1.12	1.4′0	1.35	0.55	**2.90**
20p13	21	**2.70**	1.14	1.69	1.32	0.53	2.32
21q21	23	1.40	0.28	0.56	**2.25**	0.46	1.36
Xq25	131	1.48	0.00	2.03	1.12	1.00	**2.69**

5.3 Combined analysis of families from the United States

5.3.1 Stratification of families

Families were collected from completed genotyping projects performed at the National Heart Lung and Blood Institute's Mammalian Genotyping Service (Marshfield, WI, USA) and originally ascertained for type 2 diabetes, elevated blood pressure phenotypes, systemic lupus erythematosus and obstructive sleep apnea. In total we had genome-wide genotype and phenotype information for 9,371 individuals from 3,032 families available to us. In addition to combining all available data in order to maximize the sample size we also analyzed the African-American and European-American families separately in order to reduce possible heterogeneity due to genetic and environmental variation. The relative contributions of African-Americans and European-Americans were 4,466 and 4,905 individuals from 1,628 and 1,404 families, respectively.

5.3.2 Heritability estimates

The covariate-adjusted (sex, age and cohort identifier) heritability estimates for stature were 0.87 in the African-American and 0.83 in the European-American cohorts. The heritability estimates by cohort and sex are displayed in Table 13. In the combined analysis of the African-American and European-American families we also added a covariate distinguishing the two cohorts.

Table 13. ***Heritability estimates for stature from the US families.***

	Males	Females	Combined
African-American	0.94	0.97	0.87
European-American	0.81	0.92	0.83
All	0.87	0.95	0.85

5.3.3 Genome-wide linkage results

The genome-wide linkage results by cohort and sex are shown in Figure 30. In summary, we found evidence for linkage to stature in previously reported loci on 11q23, 12q12, 15q25 and 18q23 as well as 15q26 and 19q13 which have not been linked to stature previously. The LOD scores for each locus are shown in Table 14 by cohort and sex.

Most of the linkage evidence emerged from the European-American cohort where multipoint LOD scores ≥ 2.0 were found at loci 11q23, 12q12, 15q25 and 15q26. When African-Americans were included in the analyses linkage evidence at these loci decreased suggesting some locus heterogeneity between these samples. The African-American sample did not yield any LOD scores ≥ 2.0 when analyzed separately, however their inclusion in the joint analyses increased evidence for linkage substantially at 18q23 and 19q13 suggesting common loci influending stature between the African-American and European-American cohorts.

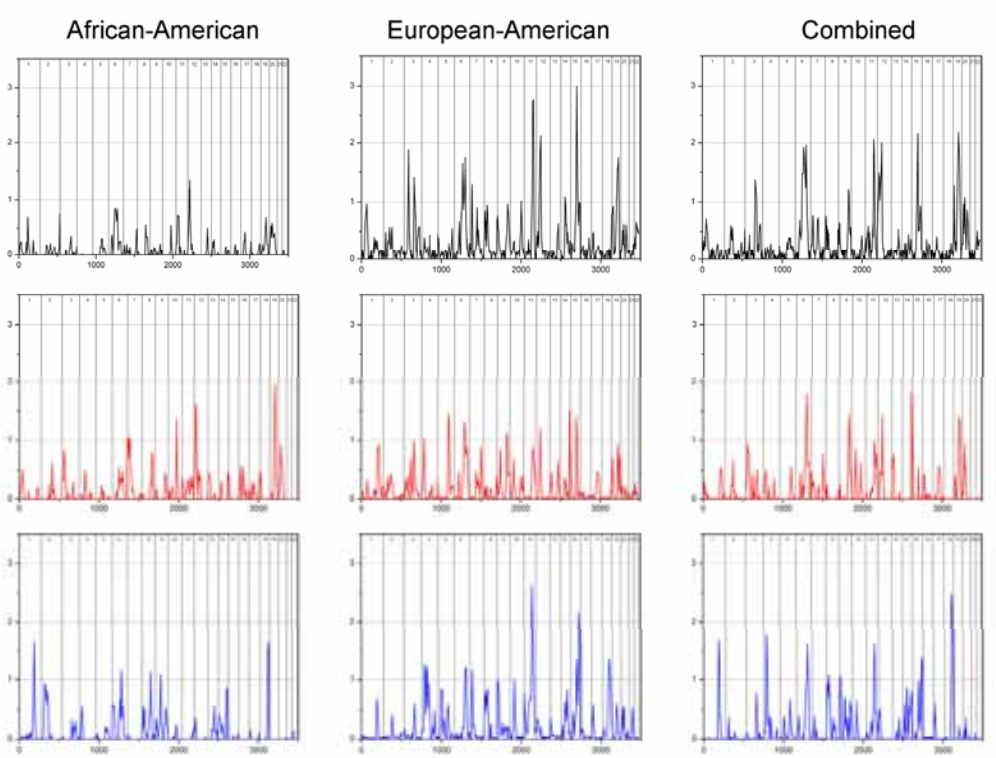

Figure 30. ***Genome-wide multipoint linkage results for stature with LOD scores on the y-axis and cumulative cM on x-axis. The left panel displays results for African-American families, the middle panel for European-American and the right panel results for all families combined. The results for males and females jointly, females and males are shown from top to bottom, respectively***

Table 14. ***Multipoint LOD scores ≥ 2.0 observed in the US families***

Locus	Peak Marker	cM	Multipoint LOD Score								
			African-American			European-American			All		
			Males	Females	Combined	Males	Females	Combined	Males	Females	Combined
11q23	D11S1986	110	0.00	0.40	0.09	2.21	0.71	**2.74**	1.35	1.00	2.06
12q12	D12S1301	58	0.00	0.49	0.22	0.22	1.19	**2.07**	0.02	1.45	2.00
15q25	D15S655	92	0.00	0.00	0.00	1.35	1.40	**3.00**	0.99	0.55	2.16
15q26	D15S1515	122	0.00	0.00	0.05	**2.15**	0.12	1.00	1.20	0.00	0.92
18q23	D18S844	133	1.54	0.00	0.17	1.01	0.03	0.05	**2.49**	0.00	0.30
19q13	D19S246	83	0.00	1.95	0.69	0.37	0.08	1.56	0.14	1.41	**2.18**

6 DISCUSSION

6.1 Heritability estimates for height

The heritability estimates show remarkable consistency across the populations in this study as is shown in Table 15. These results further establish that the heritability of adult height in modern Western societies is well over 0.80 (Reviewed by Silventoinen 2003[202]). However, as already discussed earlier (see sections 2.3.4 and 2.4.1) this may only reflect the fact that the populations included in this study are highly similar in respect to environmental influences affecting growth. Also since the samples sizes from each population are large, sampling variation is negligible and the heritability estimates are robust.

The sex-specific heritability estimates are substantially higher than the estimates from the joint analyses which may suggest the presence of sex-specific genetic elements i.e. that sex is not merely an environmental influence but also interacts with genes that influence height; therefore sex is a confounder. The heritability estimates for females are also consistently larger than those for males suggesting that environmental influences may be less relevant for females. This may suggest that females are better buffered against environmental influences on growth which has been reported previously[203-205] although not all studies agree [206].

Table 15. ***Heritability estimates from this study by sex and population***

	Males		Females		Combined	
	h^2	N	h^2	N	h^2	N
Finland	0.95	659	0.99	758	0.84	1,417
Europe	0.93	2,774	0.98	5,676	0.82	8,450
United States	0.87	3,550	0.95	5,756	0.85	9,306

6.2 Stature loci identified in the Finnish families

6.2.1 Locus 1p21

Epidemiologists have a long tradition of stratifying samples according to sex, but this approach has been relatively uninvestigated in genetic mapping studies until recent years[8,193,196,207,208]. Since sex-specific hormonal environments may influence the expressivity of genes that influence height we decided to analyze males and females separately in order to increase the statistical power to detect these genes and consequently we were able to show strong evidence for male-specific linkage to 1p21. This locus is the single strongest finding in this study (I-III).

Evidence for linkage to this locus was also reported previously in a sample of African-Americans[126], although the investigators did not perform sex-limited analyses for height. The 1-LOD drop support region (the genomic region surrounding the linkage peak where the difference in LOD score is less than 1 unit) surrounding this linkage peak harbours three genes that we considered as interesting functional candidate genes for height: EXTL2, COL11A1 and CHI3L2.

EXTL2 is an enzyme that is involved in the biosynthesis of heparin/heparin-sulfate and belongs to the hereditary multiple exostoses (HME) family of tumor suppressors. Mutations in two other members of this gene family, EXT1 and EXT2 have been shown to cause HME types I or II where approximately 50% of patients exhibit short stature. Interestingly there is also a significant excess of male patients (104:76) and male probands (28:15) in HME implying sex-specific genetic effects for this gene family.

COL11A1 encodes for one of the two alpha chains of a minor fibrillar collagen, type II collagen and is mostly expressed in cartilage tissue. Mutations in COL11A1 have been shown to cause type II Stickler syndrome[209] and Marshall Syndrome[210] that both include skeletal manifestations. The link between COL11A1 and stature is especially pronounced in Marshall Syndrome where patients are short in stature relative to unaffected family members and have a stocky build. The role of COL11A1 in growth and stature is further strengthened by the phenotype of COL11A1 null mice[211]. The null mutation produces an autosomal recessive lethal phenotype with long bones that are half the normal length and wider at the metaphyses (portion of the long bones between the epiphyses and the diaphyses).

CHI3L2 is a chondrocyte protein that is highly expressed in articular cartilage chondrocytes and has been shown as a marker of chodrocyte expansion[212] relating this protein with growth of long bones. However, this gene is relatively poorly characterized to date.

6.2.2 Family based association analysis of 1p21 in Finnish families and replication in a Finnish population cohort

We further examined the locus on 1p21 by genotyping 42 SNP markers within the positional candidate genes on 1p21 in a subsample of individuals (54 families) where samples were readily available for genotyping. These families were also extended with additional family members. In addition we included 263 individuals from 38 independent families which were not included in study I. Figure 32 gives a schematic representation of the sample used. In total this sample consisted of 874 (431 males and 443 females) individuals from 92 families with genotype data and

phenotype information. The family-based association analyses were performed using MENDEL[213]. The association analyses in the population sample were conducted with SPSS using analysis of covariance (ANCOVA).

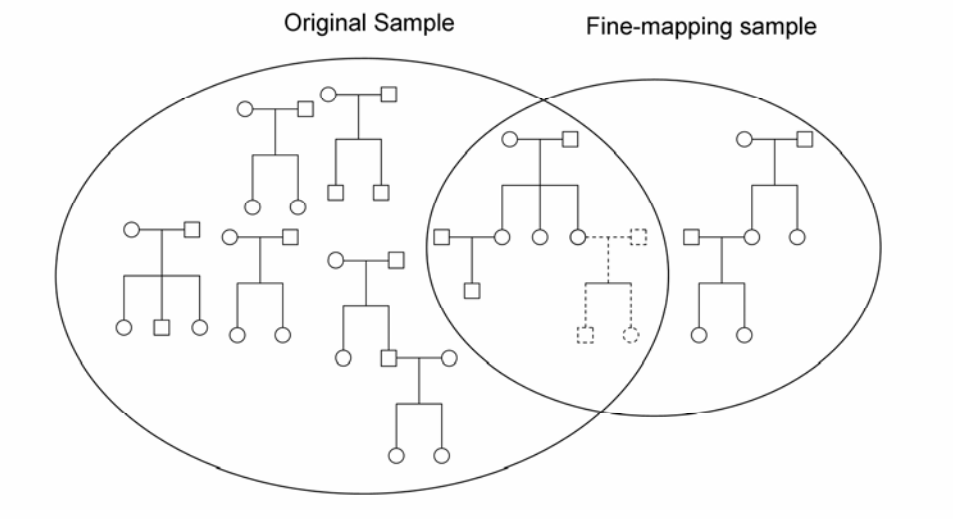

Figure 31. ***Samples used in the finemapping of the locus on 1p21. In total 54 families from the sample used in study I were included and also extended with additional family members (dotted individuals). In addition we included 38 independent families which were not included in Study I.***

Using this sample we found evidence for male-specific association (p=0.02) in a nonsynonomous SNP Pro1535Ser within COL11A1. Stratifying our sample to only families that were linked to this locus (n=45) increased the association (p=0.003). We also genotyped this SNP in a representative Finnish population cohort Health 2000 (http://www.ktl.fi/health2000) consisting of 6,542 unrelated individuals (3,023 males and 3,499 females) and replicated this association (p=0.03). In the Health 2000-cohort the effect was more pronounced in males where Ser/Ser-males were on average 1.1 cm taller compared to Pro/Ser and Pro/Pro-males, while in females they were only 0.6 cm taller (Kettunen et al., unpublished). The proportion of variance explained by this variant was estimated to be 0.1% in males, a value which is comparable to the effect of the recently reported HMGA2 variants in the GWA study by Weedon et al.[130].

6.2.3 Other loci identified in Finnish families

The locus on 4q35 has not been previously linked to stature in independent studies. Hirschhorn et al.[134] reported linkage at marker D4S1540 (LOD=1.73) in their Finnish cohort and at markers D4S3051-D4S426 (LOD=1.89) in their Botnia cohort. However, this report contained overlapping samples with Study I (both Botnia and the Finnish cohort were included in Study I) and can not therefore be considered as independent. In Study I we were able to further increase confidence for linkage at the latter locus (LOD= 2.18). The linkage region is close to the telomeric region of the long arm of chromosome four and does not harbor any obvious candidate genes for height.

The male-specific linkage observed on 9p24 is also unique to this study and is quite close to the telomeric regions which are typically quite gene-poor. Interestingly though, Mukhopadhyay et al.[193] who performed sex-limited analyses in a comparable manner to ours reported a male-specific locus (LOD=1.65 near marker D9S319) ≈ 40 cM from our linkage peak, however, due to the inherent variation in location estimates from genome-wide screens[214] it is difficult to establish whether these findings indicate the same or different loci. An interesting candidate gene, natriuretic peptide receptor B precursor NPR2 is located ≈ 5 Mb downstream of the peak marker reported by Mukhopadhyay et al.[193]. Loss-of-function mutations in NPR2 are the cause of acromesomelic dysplasia Maroteaux type (AMDM), where skeletal growth falls off sharply after birth and results in adult heights that are >5 SDs below the mean[215]. Also heterozygous individuals, though unaffected for AMDM, have been shown to be shorter than matched controls and consequently it has been suggested that 1 in 30 cases of idiopathic short stature may be explained by AMDM mutations[216].

13q12 showed linkage (LOD=2.66 at D13S221) in the females-only analyses and has been previously reported for modest linkage (LOD=1.01 at markers D13S221–GGAA29H03) by Hirschhorn et al. in a Finnish cohort[134] which was also included in this study. However, Hirschhorn et al. analyzed only 388 females compared to 758 females analyzed in this study and also did not perform sex-limited analyses which may explain why in this study the evidence for linkage was significantly stronger. In the close proximity of this locus there are two genes which when disrupted cause severe disorders where growth in stature is also affected: beta 3-glycosyltransferase-like B3GALTL and spartin SPG20. Mutations in B3GALTL are the cause of Peters-plus syndrome which is characterized by anterior eye-chamber abnormalities, disproportionate short stature, developmental delay, characteristic craniofacial features, cleft lip and/or palate[217]. Loss-of-function mutations in SPG20 on the other hand causes Troyer syndrome that is a form of spastic paraplegia with distal muscle

wasting that usually debilitates patients in early childhood. Short stature is also a prominent feature of Peters-plus patients[218].

The male-specific locus on 18q21 (LOD=2.39 at marker D18S60) overlaps well with the locus reported by Mukhopadhyay[143] (LOD=2.26 at marker D18S364) with only 5 cM separating the peak markers between these studies. However, they observed this linkage in the joint analysis of males and females and did not perform sex-limited analyses. The immediate proximity of this locus does not seem to harbor any obvious candidate genes for height.

The locus on 22q13 (LOD=2.85 at marker D22S285) has been implicated by another genome-wide screen by Ellis[135] who reported linkage at approximately 10 cM from our finding at 38 cM (LOD=2.02). One plausible candidate gene for stature in this region is adenylosuccinate lyase (ADSL) since it has been reported that some children with ADSL deficiency exhibit growth retardation[219]. However, ADSL deficiency causes severe phenotypic manifestations such as autistic features, epilepsy and mental retardation and it is unclear how this deficiency relates to normal growth in stature.

6.3 Stature loci identified in Australian and European families

6.3.1 Locus 8q21

The strongest evidence for linkage to stature in Study II was observed on 8q21 (LOD=3.28 at 97 cM) using all available individuals. Chromosome 8 has been implicated as harbouring stature QTL previously by Hirschhorn[134], although their linkage peak (LOD=2.52 at marker D8S1100) is located ≈ 50 cM downstream on 8q24. The 8q21-q24 region contains three interesting candidate genes for stature: nibrin NBN, a zinc finger transcription factor TRPS1 and exostosin EXT1. Mutations in NBN cause Nijmegen breakage syndrome which is characterized by microcephaly, growth retardation, immunodeficiency, and cancer predisposition[220]. Different mutations in TRPS1 cause both tricho-rhino-phalangeal syndrome types I and III that are both characterized by craniofacial and skeletal abnormalities including growth retardation and resulting short stature[221]. As already mentioned EXT1 belongs to the hereditary multiple exostoses (HME) family of tumor suppressors and when mutated causes hereditary multiple exostoses type 1 which is characterized by formation of numerous cartilage-capped, benign bone tumors (osteocartilaginous exostoses or osteochondromas) that are often accompanied by skeletal deformities and short stature[222].

6.3.2 Converging evidence for 8q21-q24 from an independent genome-wide association analysis

We also performed a genome-wide association scan using ≈ 317,000 SNP markers in 1,552 monozygotic female twin pairs derived from the GenomEUtwin consortium cohorts (Kettunen et al., unpublished). The twin pairs were from Australia (n=480), Finland (n=141), Denmark (n=172), Sweden (n=302) and the United Kingdom (n=457) and did not overlap with the families that were used in the genome-wide linkage screens in dizygotic twins (Study II). For each twin pair, we computed the mean height and corrected it for age and cohort effects and performed regression-based association analyses using PLINK[223]. The genome-wide association results are shown in Figure 33. Interestingly, two of the associated SNPs were on 8q24, rs1464241 ($p=4.1x10^{-5}$, rank 11.) and rs7830584 ($p=9x10^{-8}$, rank 1.) overlapping the linkage peak on 8q21. Neither of these SNPs has known regulatory functions nor is located within known genes and therefore it is unclear if these SNPs have a role in the determination of stature or if they are in linkage disequilbrium with the actual QTN(s). However, these converging findings from independent genome-wide linkage and genome-wide association screens warrant closer examination of the involvement of the 8q21-q24 locus in growth and adult height.

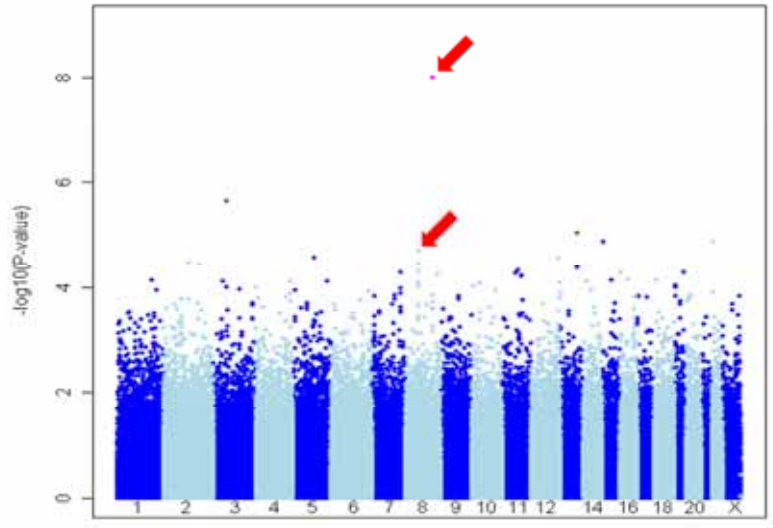

Figure 32. ***Overview of the results from the MZ GWA study for stature (Kettunen et al., unpublished). The best associated (1.) SNP rs7830584 and another top-ranking (11.) SNP rs1464241 on 8q24 overlapping the linkage signal are indicated with red arrows.***

6.3.3 Other loci identified in Australian and European families

The telomeric end of the long arm of chromosome 7 is one of the best established putative loci containing stature QTL (see Figure 31). However, in Study II we found evidence for linkage in the telomeric end of the short arm on 7p22 (LOD=2.03 at 1 cM). The linkage evidence at this locus is fairly modest and is seen only in all available individuals when males and females are analyzed jointly (Table 12). One of the prime candidate genes for stature, the growth hormone-releasing hormone receptor (GHRHR) is located near this linkage peak on 7p15. Loss-of-function mutations in GHRHR have been shown to cause isolated growth hormone deficiency also known as dwarfism of Sindh which is mainly characterized by extremely short proportionate stature and patients who are virtually devoid of other features typically associated with other types of dwarfism[147]. However, the association of common variation within this gene and short or tall stature was examined by Lettre et al. and they did not find significant evidence for any association[157].

The linkage region on 20p13 (LOD=2.90 at 15 cM and LOD=2.70 ant 21 cM) is very wide and poses some challenges when it comes to interpretation. The region could possibly contain multiple linked QTL or may contain very little recombination events in our sample due to stochastic factors. These loci map close to the linkage peak reported by Thompson et al. in Pima Indians (stature p=0.0001 and leg length p=0.001 at markers D20S66-D20S98) who also analyzed a positional candidate gene in this region, bone morphogenetic protein 2[145], which is a member of the transforming growth factor-beta (TGFB) superfamily. However, Thompson et al. did not find evidence for association to stature or leg length and the potential role of BMP2 in growth is yet to be determined although it has been reported to be capable of inducing the formation of cartilage required for bone growth *in vivo*[224].

The seemingly male-specific locus linked to stature on 21q21 has not been reported to contain stature QTL previously and does not seem to contain potential stature genes. Unlike most published genome-wide linkage screens for stature (16 out of 17) in this study we had also genotypes for X-chromosomal markers and found evidence for linkage to Xq25 (LOD=2.69 at 131 cM). Interestingly, like at the male-specific locus on 21q21, linkage evidence was increased when we restricted our analyses to dizygotic twins in order to reduce environmental variance. Liu et al.[125] showed evidence for linkage at 122 cM (two-point LOD=5.36 at marker GATA165B12P) which overlaps well with our finding. However, since they did not perform multipoint analyses, comparison between our findings is not straightforward. This region contains several genes that are associated to various syndromes characterized by short stature, including SLC6A8 mutations with X-linked creatine deficiency syndrome[225], TAZ mutations with Barth syndrome[226], CULB4 gene mutations with

MRXSC syndrome[227], SOX3 mutations with MRGH[228,229] and the MRSS syndrome locus[230] where the defective gene is unknown. Another highly interesting gene in this region is glypican 3 GPC3 that when deleted, causes Simpson-Golabi-Behmel syndrome type 1 which includes severe pre- and postnatal overgrowth leading to gigantism along with many other symptoms such as coarse facial features and congenital heart defects[231].

6.4 Stature loci identified in families from the United States

6.4.1 Locus 15q25

The strongest evidence for linkage in the US families was observed on 15q25 which shows linkage only in European-American families and is contributed by both females and males. This locus has also been previously reported with linkage to stature in the Australian sample (LOD score of 3.43 at 79 cM) included in study II. However, a seemingly distinct male-specific linkage on 15q26 was also observed in the European-American cohort. The region of these linkages contains several genes that may be relevant for stature with aggrecan 1 (AGC1) and insulin-like growth factor I receptor (IGF1R) being the most noteworthy. Aggrecan 1 has been shown to cause Kimberly type spondyloepimetaphyseal dysplasia[232] which is characterized by proportionate short stature, stocky habitus and progressive osteoarthropathy of the weight-bearing joints. There is some evidence that mutations in the IGF1R gene, resulting in IGF1 resistance, may underlie some cases of prenatal and postnatal growth failure[155].

6.4.2 Locus 12q12

Both African-American and European-American cohorts show evidence for linkage on chromosome 12 but the peak locations are slightly different. This linkage is also mostly contributed by females; however, inclusion of males adds linkage evidence suggesting a lack of sex-specific genetic effects. This locus has also been previously reported to contain a QTL for stature in a Finnish cohort[134] and contains several interesting candidate genes for stature such as SRY-box 5 (SOX5), vitamin D receptor (VDR) and collagen type 2, alpha-1 gene (COL2A1).

SOX5 has been shown to be an enhancer of chondroblast functions, controlling both the expression of extracellular matrix genes and cell proliferation[233] and to participate in COL2A1 expression activation[234]. VDR is an obvious candidate gene for stature since the vitamin D endocrine system is of paramount importance in

normal skeletal growth. VDR has also been previously linked and associated with stature[168,170]. COL2A1 encodes for type II collagen (cartilage collagen) and disruptions of its normal functions cause several forms of chondroplasias and spondyloepiphyseal dysplasia syndromes which are characterized by shortening of the trunk and limbs. Also mice carrying a partially deleted human COL2A1 gene developed the phenotype of a chondrodysplasia with dwarfism, short and thick limbs, short snout, cranial bulge, cleft palate, and delayed mineralization of bone[235].

The linkage peak on 12q12 maps close to the HMGA2 gene on 12q14 which was associated to stature in a genome-wide association study by Weedon et al.[130] thus providing converging evidence from two independent genome-wide studies utilizing complementary strategies (linkage and association). The importance of HMGA2 in growth and attainment of adult height is also supported by findings from defects in this gene that have been shown to cause gigantism in mice and humans[131,132] and dwarfism[133] in mice.

6.4.3 Locus 18q23

The locus on 18q23 is highly interesting because it showed evidence for male-specific linkage in Finnish families (Study I), US families (Study III) and in another sample of US families reported by Mukhopadhyay et al.[143]. Although Mukhopadhyay et al. used phenotypes for both sexes in their study they did model imprinting effects as well as utilized sex-specific genetic maps. In both Studies I and III the linkage observed on 18q23 is clearly sex-specific since inclusion of females dilutes evidence for linkage completely. In the US families this male-specific linkage is contributed roughly equally by both African-American and European-American cohorts although the exact peak location is slightly different between the cohorts.

The linkage peak on 18q23 lies within the region deleted in patients with chromosome 18q deletion syndrome. The phenotype of this syndrome is highly variable but commonly includes mental retardation, short stature, hypotonia, hearing impairment, and foot deformities. The likely cause of growth impairment in these patients is growth hormone insufficiency which has been associated to a deletion of an approximately 2 Mb region between markers AFM242yf2 and D18S462[236]. An excellent candidate gene for stature within this critical region is the galanin receptor (GALR1) because its substrate, galanin (GAL), is a potent stimulator of growth hormone secretion[237].

6.4.4 Other loci identified in US families

The linkage observed on 11q23 is mostly driven by European-American males which show clearly the most evidence for linkage. This locus however, may not be truly sex-specific since females also show some evidence for linkage and the total evidence for linkage is stronger when females and males are analyzed jointly. African-American families on the other hand exhibit virtually no evidence for linkage at this locus. Mukhopadhyay et al. also reported linkage peak overlapping our findings in this region[131,132,193]. This locus contains several interesting candidate genes for stature. One such gene set is a cluster of matrix metalloproteinase (MMP) genes such as MMP8, MMP10 and MMP13 which are involved in the breakdown of bone extracellular matrix by degrading proteoglycans and type I, II and III collagens. MMP13 is of particular interest since it is the likely cause of human Missouri type spondyloepimetaphyseal dysplasia[238], a skeletal disorder characterized by defective growth and modelling of the long bones and the spine.

Chromosome 19 has not been previously linked to stature in other genome-wide screens. The linked region on 19q is very large containing hundreds of known genes. The most prominent candidates are transforming growth factor-beta (TGFB1) and a cluster genes belonging to the insulin-like growth factor family of signaling molecules (IGFL1, IGFL2, IGFL3 and IGFL4). TGFB1 is highly expressed in developing cartilage, endochondral and membrane bone, and skin, suggesting a role in the growth and differentiation of these tissues and it also regulates the actions of many other growth factors. The physiologic functions of IGF-like proteins are not well defined but given the homology with IGF proteins that have central roles in growth and development (such as IGF1 and IGF2), it is likely that genes of the IGF-like family act as growth modulators as well[239].

6.5 Stature loci localized in this study compared to previous findings

6.5.1 The stature gene map

Including this study, there have been 17 genome-wide linkage screens and one genome-wide association screen that have reported linkage findings in every chromosome except chromosomes 10, 16, 19 and Y. However in Study III we report evidence for a QTL on chromosome 19 as well. To facilitate the accumulation of genetic evidence in stature gene mapping we have established an Internet knowledge resource The Stature Gene Map (www.genomeutwin.org/stature_gene_map.htm) that summarizes all published linkage findings for stature.

The schematic diagram from the web site (Figure 33) visualizes the current knowledge on QTL underlying height. From this figure it is clear that most findings are from single reports, some loci are supported by two and only the loci on 6q25, 7q35-36 and 9q21-22 by three studies.

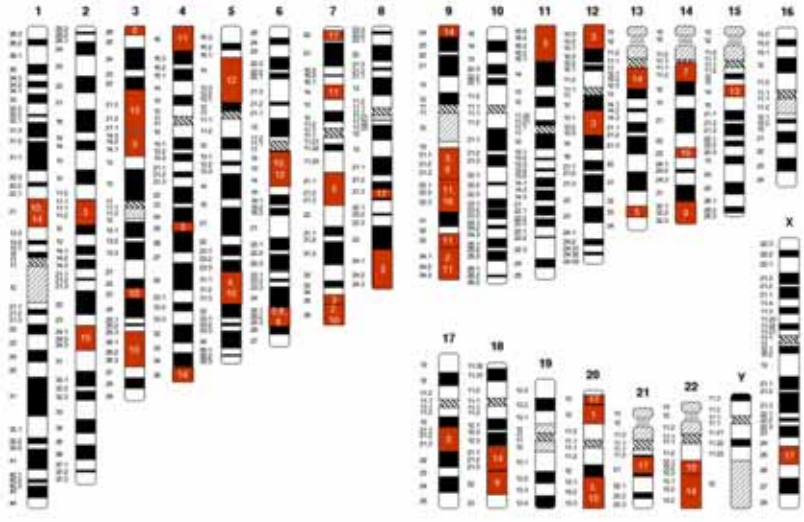

Figure 33. ***Overview of the results from published genome-wide QTL screens for stature where multipoint LOD scores over 2.0 have been reported. The numbers refer to the original publications that are accessible via hyperlinks at the Stature Gene Map web site. Study I is shown with number 14 and study II with 17. Study III is not shown here because it has not been published yet. http://www.genomeutwin.org/stature_gene_map.htm***

6.5.2 Replication in genome-wide studies

Naturally the most convincing loci are those reported by multiple studies; however, converging evidence from multiple linkage studies may be complicated because the support regions for linkage peaks are typically very large, spanning even tens of centimorgans and hundreds of genes and maximum likelihood estimates of QTL locations are subject to chance variation[214]. In QTL analysis it has also been shown that the precision of the location estimate is also positively correlated with the QTL effect[240]. The QTL effect on the other hand may vary between studies due to many factors; for example due to sampling variation or differences in genetic and environmental influences. Thus it is difficult to distinguish between overlapping and independent findings when they are located, for example, on adjacent chromosome bands.

There is also considerable debate regarding the statistical criterion for declaring replication of a previously reported linkage signal although some have postulated thresholds for replication[241-243]. A common confusion when investigators claim replication is the lack of multiple testing corrections. For example, authors who find a LOD score in their genome-wide scan with a nominal p-value of 0.05 that overlaps with a previous genome-wide significant linkage report and claim replication thus ignoring the multiple testing they have performed in their genome-wide scan. Statististical significance is always a subjective decision and thresholds such as $p \leq 0.05$ or $LOD \geq 3.0$ are completely arbitrary. It is clear that most test statistics are dependent on many features such as family structures and allele frequencies and therefore empirical significance might be a more useful method of comparing statistical findings from multiple independent studies.

6.6 Conclusions from loci discovered in this study

The loci discovered in this study well represent the findings from previous genome-wide screening of stature QTL: most of the discovered loci are unique to a single study while some findings overlap between studies. However, some of the loci implicated in this study merit more focus than others.

The male-specific locus on 1p21 identified in the Finnish families (Study I) seems particularly interesting since we were able to show that a specific variant Pro1535Ser in the positional candidate gene COL11A1 was associated with stature in Finnish families where the initial linkage was established and was also replicated in a large population cohort that is representative of the Finnish general population. Using this population cohort we estimated that this variant explains 0.1% of the population variance in height in males. To our knowledge, this variant is the first sex-specific quantitative trait nucleotide (QTN) discovered that is convincingly

associated with stature in the general population (Kettunen et al., unpublished). In addition, it is possible that this locus has an effect on stature in other populations as well, since it has been reported for linkage in US African-Americans as well[126].

The locus on 8p21 linked to stature in Australian and European families (Study II) is also of special interest since there is converging evidence from our genome-wide association screen in MZ twins that specific variants at this locus may influence stature. We found two variants that were highly associated with stature in an independent GWA screen (Kettunen et al. unpublished) and we are currently investigating this locus further.

There is also a highly interesting overlap between the 12q21 locus linked to stature in our study (Study III) and the recently reported association of HMGA2 to stature from the first published genome-wide association study for stature[130]. Unfortunately our data does not allow us to further dissect whether this gene explains the observed linkage in our study or if there are other stature QTL at this locus.

Furthermore, in this study we were able to show converging linkage evidence for the 18q21-q23 region from the Finnish and the US families (Studies I and III) which was prior to this study an "orphan" stature locus where only one study had reported linkage to stature[143]. Now, this locus joins the group of prime candidate loci 6q25, 7q35-q36 and 9q21-q22 that are reported for linkage to stature by three independent genome-wide linkage screens.

Also, our results imply that many loci influencing height are sex-dependent - in this study 8 out of the 16 loci that showed evidence for linkage (LOD $\geq$ 2.0) were either exclusive for one sex or the linkage was stronger in the sex-stratified analysis. The major finding of this study the Pro1535Ser variant in COL11A1 demonstrates clearly how the effect is different in males and females where Ser-homozygote males are on average 1.1 cm taller, while females are 0.6 cm taller than their counterparts with other genotypes. It is plausible that the larger effect in males compared to females was the critical factor that allowed the detection of linkage in the males-only analyses in study I since no linkage was observed at this locus in the combined analysis of males and females nor in the females-only analyses. Until recent years the potential sex-specific architecture of human traits has been underappreciated and relatively uninvestigated in gene mapping studies, where sex has usually been treated as a simple environmental covariate. Reports of sex-specific or sex-dependent genetic findings have now begun to surface [8,193,196,207,208] and probably will do so even more in the future.

6.7 The importance of studying stature

Understanding the genetics of human height is important for several reasons. First, linear growth is one of the fundamental characteristics of childhood, and deviation from normal patterns of growth is a common cause of medical evaluation and referral to pediatric clinics. Also, because the height of an individual may have important socioeconomic and psychosocial implications there is special interest in the height of children of shorter than average parents.

Secondly, epidemiological studies have shown many associations between height and general health indicators as well as between height and several specific diseases such as cancers and cardiovascular diseases. Although these associations may be examples of classical confounding, where two variables correlate due to a third shared causative factor, there may indeed be shared etiology as well and unraveling the genetics of height may give insight to these etiologies.

Thirdly, even though there is extensive knowledge on hormonal regulation of growth we still lack a basic understanding of biological mechanisms of growth and the factors that produce observed variation in growth tempo and potential. The study of height may shed light on these basic biological mechanisms as well.

Lastly, height may serve as a model phenotype of a polygenic trait study which may yield important lessons for methodology and study design of other polygenic phenotypes of interest.

6.8 Lessons from genome-wide studies of stature

6.8.1 Heritability and mappability

By definition, heritability is an aggregate measure that includes the joint effect of all genes that influence the trait and it does not give any insight to the number genes affecting the trait or to the relative proportions of variance explained by each gene. In general, traits for which heritability is low are not amenable to QTL mapping[244]. However, this does not translate to the fact that traits with high heritabilities are amenable to successful QTL mapping either. QTL mapping is most likely to succeed for so called *oligogenic* quantitative traits that are controlled by only a few bi-allelic loci that are largely responsible for contributing genotypic variation to variation in the phenotype (major genes) and an indefinite number of loci with small effects[245,246]. For truly *polygenic* traits, which are controlled a large number of QTL with small and roughly equal effects (i.e. no major genes), locating those QTL with current mapping methods may be impossible[247]. Accumulating knowledge from this and other studies

suggest that stature may be indeed highly polygenic, since only a few loci have been consistently mapped and the identified variants explain only minute proportions of the observed variance despite the very large samples in some studies.

Another concern is the accumulating evidence for the highly polygenic background of quantitative traits. For height, the identified variants on HMGA2 and COL11A1 explain only 0.3% and 0.1% of the observed variation suggesting that there may be hundreds or even thoudands of QTL of similar or even smaller effects. Visscher et al.[248] estimated by genome partitioning in a sample of 11,214 sibling pairs that all autosomes and the X chromosome contribute to the variation in height. They also described that in the "best-case" scenario the average QTL would explain 3.9% of the phenotypic variance (90% heritability equally divided into 23 chromosomes each harboring a single QTL) and that even in this case a linkage study of 57,830 sib pairs would be required for 80% power with an alpha level of 0.0001. Therefore even the largest studies such as this one would be likely to be underpowered. In the study of model organisms it is well established that the effects (QTL specific heritabilities) of identified QTL are distributed in an extreme L-shaped manner, where a few loci have relatively large effects, more loci have moderate and may have very small effects[249-255]. However, the effects are typically less than 5%[256]. It should be noted though that the observed L-shaped distribution may also arise from several statistical artefacts as well[257].

6.8.2 Maximizing sample size by combining data

Common explanation for failure to identify loci in genetic mapping is a lack of statistical power due to small sample sizes[258]. One of the key aims of this study was to examine the usefulness of combining data across multiple primary sources. Although it was clear that combining data is also likely to increase heterogeneity due to genetic and environmental sources, our hypothesis was that major loci containing ancient variants that are shared across diverse populations may be detectable given large enough sample size. However, if stature is controlled only by QTL of small effects (e.g. < 5% of the observed phenotypic variance) even this study is likely to be underpowered. For most loci, we could definitely demonstrate the value of combining data since they were not detectable in individual subsamples. For other loci, however we noticed that reducing variation by stratification was essential for locus identification. Stratification may also produce false positives and false negatives due to sampling variation if sample size is small. Still if QTL effects are very small as some model organism data and recent data from human studies suggests even this study may have been underpowered to detect most QTL underlying height and one requires luck to discover true QTL.

6.8.3 Information content

Another potential factor that may explain the limited success in genome-wide linkage studies is the relatively low information content due to missing founder genotypes and relatively sparse marker maps of traditional genome-wide screens. In a systematic analysis of 101 genome-wide linkage scans[258] the average marker density was 11 cM which is far too sparse in order to extract all inheritance information[259] especially if there are missing founder genotypes in the data. Since the expected LOD score and therefore the power of linkage analysis is proportional to the inheritance information extracted from the marker data[260] it is likely that most genome-wide linkage studies have been seriously underpowered. Many empirical [261] and simulation studies have shown that regenotyping families that have already been collected with high-density SNP maps is highly beneficial. Since sample ascertainment and collection are very time-consuming and expensive stages in genetic studies this would very likely be worth the effort.

6.8.4 Stratification strategies

It has become increasingly evident from this and other studies that special ascertainment schemes may be essential for dissection of the genetic architecture of oligo- and polygenic traits. This study has demonstrated the usefulness of reducing environmental variation due to sexual dimorphism and unshared familial environment by analyzing males and females (who differ greatly in hormonal environment) separately and by sampling dizygotic twins who share fetal, childhood and adolescent influences on growth and attainment of adult height. Also, when there is population stratification, such as in the US families, in most cases analyzing ethnic groups separately results in greater statistical power even though sample size is reduced.

Reducing genetic and environmental heterogeneity is clearly an efficient strategy in genetic mapping and may also be a more critical determinant than robust statistical power based on sample size. This is evident from the results from this study; the most convincing locus finding in our genome-wide studies was from the smallest and the most homogeneous Finnish study sample - a finding which was also recently replicated in an independent population cohort (Kettunen, unpublished). This should not be surprising since multiple successful gene identifications have been performed utilizing homogeneous isolated or relatively isolated populations (see Peltonen et al. [262] and De La Chapelle[263] for review). Also, systematic analyses have shown that the only two critical factors determining successful locus identification are sufficiently large sample size and genetically and environmentally homogeneous study

sample[258]. Results from this study suggest that given sufficient sample size the latter may be a more useful strategy.

6.9 Future strategies for studying stature

6.9.1 Sample population

This study has demonstrated the importance of sample homogeneity over sample size and no matter which study designs are utilized sample choice is of paramount importance in any genetic study. Ideally the sample should be homogeneous in terms of environmental and genetic variation because 1) environmental influences add random noise to genetic analyses thus reducing statistical power and because 2) all genetic mapping methods are susceptible to false positive and false negative findings in the presence of locus and allelic heterogeneity (although association analyses are more susceptible in general[247] whilst linkage may succeed[264]). The recent availability of very dense genome-wide data enables the stratification of samples based on their actual genetic resemblance[265,266] instead of relying on geographic location, genealogy or self-reported ancestry and will provide an excellent tool for detecting population stratification. One should also bear in mind the importance of environmental homogeneity and attempt to minize all sources of heterogeneity by clever study design and careful documentation of relevant exposures.

6.9.2 Phenotype of choice

Often the phenotype of interest may be the end result of a complex multibranched pathway. However, focusing on intermediate stages (endophenotypes or intermediate phenotypes) of those pathways may allow more successful dissection of the genetic architecture of the phenotype. The use of expression levels in linkage analysis has demonstrated that this is a powerful approach[267-270].

For example height is an aggregate trait that is composed mainly of lengths of the long bones, the spine and the cranium. It is not known whether these genes are controlled by separate or shared sets of genes or a mixture of both. If there are genes that influence the growth of specific components of height, performing genetic analyses for these components separately is bound to improve the power to detect underlying QTL.

One should also bear in mind that growth is a longitudinal process that occurs in fetal life, childhood and adolescence. However, genetic studies of stature are cross-sectional and focus only on adult height which is the endpoint of this growth

process. It is well known that growth rate in relative and absolute terms varies greatly between these developmental stages and they may also be under specific genetic control. Restricting analyses to these specific growth periods could permit the genetic dissection of the growth process in more detail. Also individuals vary greatly in growth velocity within specific stages which may also be genetic in origin; focusing on these stages might unravel these genetic differences as well.

Another challenge is how to better integrate the knowledge accumulated by the scientific community in genetic analyses. For example there is extensive knowledge on environmental factors that influence growth and subsequent adult height yet this data is rarely collected and incorporated into the ascertainment schemes and genetic models. In addition molecular and statistical geneticists should collaborate more with auxologists, endocrinolgists and anthropologists especially in the design phase of the study to better take into account all relevant variables and other considerations. If these data could be included in an appropriate manner they could accelerate genetic discoveries immensly

6.9.3 Study design and analysis

There have been many genome-wide screens for stature QTL, but unfortunately these efforts have been characterized by lack of consistency between studies and surprisingly low statistical evidence for QTL despite large samples (such as the ones included in this study). Genome-wide association studies utilizing large numbers of unrelated individuals have recently been proposed as a method of choice for identifying genes underlying polygenic traits[271,272]. Unfortunately, for many investigators, the GWA approach seems so appealing that they may not bother with collecting family samples since it is more time-consuming than collecting unrelated individuals. The arguments for the recent frenzy for GWA studies are both practical as well as statistical; it is far less time consuming and expensive to collect unrelated individuals as opposed to families and given that correct markers (i.e. the quantitative trait nucleotides (QTN) affecting the trait or markers in strong linkage disequilibrium with them) are genotyped and the genetic mechanism is fairly simple (e.g. only a few common variants) association analysis is more powerful than linkage analysis[264,273].

Table 16. ***Simplistic comparison between relevant differences in genome-wide linkage and genome-wide association designs.***

Linkage	Association
Requires family samples	Can be performed with unrelated individuals or family samples
Low resolution since it focuses on observed recombinations	High resolution because it focuses on historical recombinations
Immune to allelic heterogeneity	Sensitive to allelic heterogeneity
Immune to false positives due to population stratification	Sensitive to false positives due to population stratification (in population samples)
Optimal for rare alleles of moderate-to-large effect	Optimal for common alleles and can detect modest effects

One should remember that family-based linkage mapping and association mapping in unrelated individuals are optimal under very different genetic models (Table 16) and therefore it is unwise to invest solely in one or the other since we do not know a priori the genetic architechture of the trait we are interested in. In the case of rare monogenic disease, multiple rare variants at linked loci (allelic heterogeneity) seem to be the rule not the exception[245,274]. For common polygenic disease and quantitative traits this question is still unanswered - there are examples for both common[275] and rare alleles[276] and theoretical and empirical studies suggest a role for both rare and common variants[277,278]. The power of GWA studies relies heavily on linkage disequilibrium since they are unlikely to genotype the actual QTN[264] but rely on LD between the typed variants and the actual QTN. Finally, also the extravagant multiple testing involved in GWA studies reduces statistical power drastically due to the requirement for stringent significance thresholds.

One complementary approach of combining family-based linkage and GWA analysis was performed recently by Cheung et al.[279] who utilized the Centre d'Etude Polymorphism (CEPH) families where they initially performed a family-based linkage study for 3554 gene expression phenotypes[268]. For the strongest linkages (N=27) they first performed regional association and later genome-wide association using the founders from the same families and they were able to show concordant results using these two approaches for many of the loci (n=15). Considering the inherent qualities of family-based linkage and association analysis in unrelated individuals and empirical studies it is clear that genome-wide linkage mapping in families and genome-wide association mapping in population cohorts should rather be considered as complementary, not alternative, strategies in mapping polygenic traits.

7 CONCLUDING REMARKS

The foundation of this study has been an extensive national and international scientific collaboration that has resulted in the pooling of intellectual resources and study materials. Without such collaboration modern science would not excel at the pace we have seen in recent years.

During this study the emphasis in genetic mapping has shifted from genome-wide linkage studies in families to genome-wide association studies utilizing unrelated individuals due to the availability of sufficient SNP marker via the HapMap-project and the advent of rapid and affordable genotyping technologies.

If the majority of the genetic background of stature is explained by combinations of relatively common variants genome-wide association studies may provide novel insights into the genetics of growth and stature. However, if stature is largely determined by rare variants that have a large impact for individuals but very little impact on the population level, family-based linkage studies are better geared for identifying them. Therefore, although the genome-wide association approach holds great promise it should be considered as a complementary strategy to genome-wide linkage not as the sole method of choice in the genetic mapping of human complex traits. Also, with proper study design these two approaches need not be mutually exclusive but their power can be combined.

It does seem that considering human stature as an “easy” model trait is a fallacy, although stature inarguably possesses desirable features for genetic mapping such as easy and reliable measument of the trait phenotype. Even though we have been able to identify a few genetic variants underlying stature variation it is likely that we will need deeper phenotyping, careful sample ascertainment, insightful study design, sophisticated statistical modeling and careful consideration in accumulating knowledge to fully understand the genetics and biology of growth and its endpoint, human stature.

8 ACKNOWLEDGEMENTS

This study was carried out at the Department of Molecular Medicine of the National Public Health Institute, Helsinki. I wish to thank the former head of the institute, Professor Jussi Huttunen, the new director, Professor Pekka Puska and the head of the department of molecular medicine Adjunct professor Anu Jalanko for providing state-of-the-art research facilities that have made this work possible.

This work was financially supported by the GenomEUtwin-project, Helsinki Biomedical Graduate School and the Orion-Farmos and Sigrid Juselius Foundations.

First and foremost, I wish to express my deepest gratitude to my supervisors Academy Professor Leena Peltonen-Palotie and Adjuct Professor Markus Perola. I have been very fortunate to work with such insightful and knowledgeful scientists that possess a true passion for their work. Leena's broad expertise, immense scientific drive and achievements have been a true inspiration for me from day one. Markus, I thank you for giving me an opportunity to work in your supervision and providing such an inspiring, motivated and enjoyable work atmosphere as well as all the good times - I couldn't have asked for more.

This work has been a culmination of national and international scientific collaboration of gigantic proportions and there are just too many collaborators to thank by name. Your kindness, trust and collaborative spirit have been the foundation of this work.

I want to sincerely thank the external members of my thesis committee Professor Kimmo Kontula and Dr. Vesa Ollikainen who have given me a share of their valuable time and have supported me throughout my career as a graduate student.

I am deeply grateful to Adjuct Professors Päivi Onkamo and Pekka Uimari for reviewing my thesis - their comments were invaluable for improving the manuscript. I would also wish to thank Donald Smart for thorough language revision of the manuscript. My utmost gratitude goes to Dr. Harald Göring for accepting the role as Opponent in my thesis defense.

I wish to express my gratitude to senior scientists who have helped along the way: Marjo Kestilä, Ismo "Iski" Ulmanen, Kaisa Silander, Teppo Varilo, Iiris Hovatta, Juha Saharinen, Vesa Olkkonen and Matti Jauhiainen. Very special thanks go to Marjo whose help during the preparation of this thesis and in unraveling all the red tape was invaluable. I am greatly indebted to you Iski, for initially hiring me to work at the Department as a mere first year biology student and for your career advice that led to work for Markus. Thank you.

I would also like to acknowledge our wonderful secretaries Sari Kivikko, Sari Mustala, Tuija Svahnbäck, Sanna Tossavainen and Mika Kivimäki for helping me with countless tasks with amazing efficiency and positive attitude. Jari Raikko is also thanked for all the IT-support during the years.

I am greatly indebted to Tero Hiekkalinna who has always been there for me when I have needed assistance with anything computer-related and for his overwhelmingly positive attitude and never-ending willingness to help. I also want to thank Tero for all the good times we have had outside work – its hammer time! The contribution of Johannes Kettunen has also been vital in this work and is greatly appreciated. I am deeply greatful to Heidi Lilja and Elina Suviolahti for all their help with the Finnish families used in this study. I also want to express my gratitude for all my other friends at the office Juri, Tomi, Petri, Hannu and Teemu for the relaxed and enjoyable atmosphere that made coming to work so much more fun.

I could not think of a better lab to work in because of the terrific peers, many of you have become my friends during my years in the lab. Very special thanks go to the former and present members of our Quantitative Genetics group: Johannes, Kati, Mervi, Kirsi, Elina C, Katja, Annina, Henna, Juha K, Kaisu and Antti. I also want to specially thank Jussi, Nora, Outi, Henna, Tiina, Suvi, Mari, Tanja, Markus L, Anu, Suvi, Virpi, Anna-Maija, Heli H, Annika, Jonna, Heidi, PP, Krista, Elina M, Marika, Annina L, Pekka and Minttu for your company and all the great parties. All our wonderful lab technicians are thanked for all the smiling faces every morning. Also, I wish to thank my friends from the lab who have left over the years: Tony, Denis, Sarah, Heli K, Eeva, Suski, Jenny, Joni, Emma, Anna K, Marie, Kismat and Aimee.

My friends mean the world to me and although there are too many to name there are some that need to be mentioned: Juha, Simon, Vesku, Sami, Mikko, Kimmo, Teppo, Pertti, Masa, Apina, Juti, Jope, Teme, Stefu, Holmi, Tede, Tommi, Samuli, Cabe, Tupla-W, Junde, Make, Andu, Danny, KT, Timo, Kivistö, Masa, Fille, Leo, Jari, Bobo, Tebian, Kimmo L, Samppa and Jaakko. Thank you for all the good times.

My deepest appreciation and love goes to my family, my sister Suvi and her husband Jukka and my adorable niece Lotta and wonderful nephews Santeri and Valtteri. Mom and Dad, you will always be in my heart.

Helsinki, December 6th, 2007

Sampo Sammalisto

9 REFERENCES

1. MacGillivray MH, Morishima A, Conte F, Grumbach M, Smith EP: Pediatric endocrinology update: an overview. The essential roles of estrogens in pubertal growth, epiphyseal fusion and bone turnover: lessons from mutations in the genes for aromatase and the estrogen receptor. *Horm Res* 1998; **49 Suppl 1:** 2-8.

2. Nilsson A, Ohlsson C, Isaksson OG, Lindahl A, Isgaard J: Hormonal regulation of longitudinal bone growth. *European journal of clinical nutrition* 1994; **48 Suppl 1:** S150-158; discussion S158-160.

3. Roche AF, Sun SS: *Human Growth: Assessment and Interpretation*: Cambridgre University Press, 2003.

4. Rogol AD: Growth hormone: physiology, therapeutic use, and potential for abuse. *Exercise and sport sciences reviews* 1989; **17:** 353-377.

5. Galton F: *Natural Inheritance*: McMillan, 1889.

6. Stigler SM: Francis Galton's Account of the Invention of Correlation. *Statistical Science* 1989; **4:** 73-79.

7. Arnold SJ: Multivariate Inheritance and Evolution: A Review of Concepts; in: Boake CRB (ed): *Quantitative Genetic Studies of Behavioral Evolution*: University Of Chicago Press, 1994, 1. edition edn.

8. Pan L, Ober C, Abney M: Heritability estimation of sex-specific effects on human quantitative traits. *Genetic epidemiology* 2007; **31:** 338-347.

9. Pilia G, Chen WM, Scuteri A *et al*: Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* 2006; **2:** e132.

10. Cole TJ: Secular trends in growth. *The Proceedings of the Nutrition Society* 2000; **59:** 317-324.

11. Silventoinen K: Body height: determinants and associations with social position and adult health., Helsingin yliopisto, helsinki, 2000.

12. Brush G, Harrison GA, Zumrawi FY: A path analysis of some determinants of infant growth in Khartoum. *Annals of human biology* 1993; **20:** 381-387.

13. Vella V, Tomkins A, Borghesi A, Migliori GB, Oryem VY: Determinants of stunting and recovery from stunting in northwest Uganda. *International journal of epidemiology* 1994; **23:** 782-786.

14. Cole TJ: The secular trend in human physical growth: a biological view. *Economics and human biology* 2003; **1:** 161-168.

15. Karlberg J, Jalil F, Lam B, Low L, Yeung CY: Linear growth retardation in relation to the three phases of growth. *European journal of clinical nutrition* 1994; **48 Suppl 1:** S25-43; discussion S43-24.

16. Allen LH: Nutritional influences on linear growth: a general review. *European journal of clinical nutrition* 1994; **48 Suppl 1:** S75-89.

17. Zerfas AJ: Epidemiology and nutrition; in: Falkner F, Tanner JM (eds): *Human Growth*. New York: Plenum Press, 1986, vol 3.

18. Prentice A, Bates CJ: Adequacy of dietary mineral supply for human bone growth and mineralisation. *European journal of clinical nutrition* 1994; **48 Suppl 1:** S161-176; discussion S177.

19. Dallman PR: Iron deficiency in the weanling: a nutritional problem on the way to resolution. *Acta paediatrica Scandinavica* 1986; **323:** 59-67.

20. Johnson JA, Kumar R: Vitamin D and renal calcium transport. *Current opinion in nephrology and hypertension* 1994; **3:** 424-429.

21. Fawzi WW, Herrera MG, Willett WC, Nestel P, el Amin A, Mohamed KA: Dietary vitamin A intake in relation to child growth. *Epidemiology (Cambridge, Mass* 1997; **8:** 402-407.

22. Evain-Brion D, Porquet D, Therond P *et al*: Vitamin A deficiency and nocturnal growth hormone secretion in short children. *Lancet* 1994; **343:** 87-88.

23. Raifen R, Altman Y, Zadik Z: Vitamin A levels and growth hormone axis. *Horm Res* 1996; **46:** 279-281.

24. Allen LH: Biological mechanisms that might underlie iron's effects on fetal growth and preterm birth. *The Journal of nutrition* 2001; **131:** 581S-589S.

25. Christian P, Khatry SK, Katz J *et al*: Effects of alternative maternal micronutrient supplements on low birth weight in rural Nepal: double blind randomised community trial. *BMJ (Clinical research ed* 2003; **326:** 571.

26. Eide MG, Oyen N, Skjaerven R, Nilsen ST, Bjerkedal T, Tell GS: Size at birth and gestational age as predictors of adult height and weight. *Epidemiology (Cambridge, Mass* 2005; **16:** 175-181.

27. Kusin JA, Kardjati S, Houtkooper JM, Renqvist UH: Energy supplementation during pregnancy and postnatal growth. *Lancet* 1992; **340:** 623-626.

28. Stanner SA, Bulmer K, Andres C *et al*: Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *BMJ (Clinical research ed* 1997; **315:** 1342-1348.

29. Cornelius MD, Goldschmidt L, Day NL, Larkby C: Alcohol, tobacco and marijuana use among pregnant teenagers: 6-year follow-up of offspring growth effects. *Neurotoxicology and teratology* 2002; **24:** 703-710.

30. Rona RJ, Chinn S, Florey CD: Exposure to cigarette smoking and children's growth. *International journal of epidemiology* 1985; **14:** 402-409.

31. Shu XO, Hatch MC, Mills J, Clemens J, Susser M: Maternal smoking, alcohol drinking, caffeine consumption, and fetal growth: results from a prospective study. *Epidemiology (Cambridge, Mass* 1995; **6:** 115-120.

32. Stephensen CB: Burden of infection on growth failure. *The Journal of nutrition* 1999; **129:** 534S-538S.

33. Brown KH: Diarrhea and malnutrition. *The Journal of nutrition* 2003; **133:** 328S-332S.

34. Skerry TM: The effects of the inflammatory response on bone growth. *European journal of clinical nutrition* 1994; **48 Suppl 1:** S190-197; discussion S198.

35. Tattersall RB, Pyke DA: Growth in diabetic children. Studies in identical twins. *Lancet* 1973; **2:** 1105-1109.

36. Eveleth PB: Differences between ethnic groups in sex dimorphism of adult height. *Annals of human biology* 1975; **2:** 35-39.

37. Gaulin SJ, Boster JS: Human marriage systems and sexual dimorphism in stature. *American journal of physical anthropology* 1992; **89:** 467-475.

38. Gustafsson A, Lindenfors P: Human size evolution: no evolutionary allometric relationship between male and female stature. *Journal of human evolution* 2004; **47:** 253-266.

39. Seeman E: The structural basis of bone fragility in men. *Bone* 1999; **25:** 143-147.

40. Gatford KL, Egan AR, Clarke IJ, Owens PC: Sexual dimorphism of the somatotrophic axis. *The Journal of endocrinology* 1998; **157:** 373-389.

41. Ellis JA, Stebbing M, Harrap SB: Significant population variation in adult male height associated with the Y chromosome and the aromatase gene. *The Journal of clinical endocrinology and metabolism* 2001; **86:** 4147-4150.

42. Ogata T, Goodfellow P, Petit C, Aya M, Matsuo N: Short stature in a girl with a terminal Xp deletion distal to DXYS15: localisation of a growth gene(s) in the pseudoautosomal region. *Journal of medical genetics* 1992; **29:** 455-459.

43. Ogata T, Matsuo N: Sex chromosome aberrations and stature: deduction of the principal factors involved in the determination of adult height. *Hum Genet* 1993; **91:** 551-562.

44. Rao E, Weiss B, Fukami M *et al*: Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature genetics* 1997; **16:** 54-63.

45. Jousilahti P, Tuomilehto J, Vartiainen E, Eriksson J, Puska P: Relation of adult height to cause-specific and total mortality: a prospective follow-up study of 31,199 middle-aged men and women in Finland. *American journal of epidemiology* 2000; **151:** 1112-1120.

46. Peck AM, Vagero DH: Adult body height, self perceived health and mortality in the Swedish population. *Journal of epidemiology and community health* 1989; **43:** 380-384.

47. Forsen T, Eriksson JG, Tuomilehto J, Teramo K, Osmond C, Barker DJ: Mother's weight in pregnancy and coronary heart disease in a cohort of Finnish men: follow up study. *BMJ (Clinical research ed* 1997; **315:** 837-840.

48. Kaplan GA, Salonen JT: Socioeconomic conditions in childhood and ischaemic heart disease during middle age. *BMJ (Clinical research ed* 1990; **301:** 1121-1123.

49. Rahkonen O, Lahelma E, Huuhka M: Past or present? Childhood living conditions and current socioeconomic status as determinants of adult health. *Social science & medicine (1982)* 1997; **44:** 327-336.

50. Davey Smith G, Hart C, Upton M *et al*: Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *Journal of epidemiology and community health* 2000; **54:** 97-103.

51. Gertler MM, White PD: *Coronary heart disease in young adults: a multidisciplinary study*, Cambridge, MA: Harvard University Press, 1954.

52. McCarron P, Okasha M, McEwen J, Smith GD: Height in young adulthood and risk of death from cardiorespiratory disease: a prospective study of male former students of Glasgow University, Scotland. *American journal of epidemiology* 2002; **155:** 683-687.

53. Njolstad I, Arnesen E, Lund-Larsen PG: Body height, cardiovascular risk factors, and risk of stroke in middle-aged men and women. A 14-year follow-up of the Finnmark Study. *Circulation* 1996; **94:** 2877-2882.

54. Fisher LD, Kennedy JW, Davis KB *et al*: Association of sex, physical size, and operative mortality after coronary artery bypass in the Coronary Artery Surgery Study (CASS). *The Journal of thoracic and cardiovascular surgery* 1982; **84:** 334-341.

55. Albanes D, Jones DY, Schatzkin A, Micozzi MS, Taylor PR: Adult stature and risk of cancer. *Cancer research* 1988; **48:** 1658-1662.

56. Swanson CA, Jones DY, Schatzkin A, Brinton LA, Ziegler RG: Breast cancer risk assessed by anthropometry in the NHANES I epidemiological follow-up study. *Cancer research* 1988; **48:** 5363-5367.

57. van den Brandt PA, Dirx MJ, Ronckers CM, van den Hoogen P, Goldbohm RA: Height, weight weight change, and postmenopausal breast cancer risk: The Netherlands Cohort Study. *Cancer Causes Control* 1997; **8:** 39-47.

58. Albanes D: Height, early energy intake, and cancer. Evidence mounts for the relation of energy intake to adult malignancies. *BMJ (Clinical research ed* 1998; **317:** 1331-1332.

59. Frankel S, Gunnell DJ, Peters TJ, Maynard M, Davey Smith G: Childhood energy intake and adult mortality from cancer: the Boyd Orr Cohort Study. *BMJ (Clinical research ed* 1998; **316:** 499-504.

60. Cavelaars AE, Kunst AE, Geurts JJ *et al*: Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Annals of human biology* 2000; **27:** 407-421.

61. Power C, Manor O, Li L: Are inequalities in height underestimated by adult social position? Effects of changing social structure and height selection in a cohort study. *BMJ (Clinical research ed* 2002; **325:** 131-134.

62. Silventoinen K, Kaprio J, Lahelma E: Genetic and environmental contributions to the association between body height and educational attainment: a study of adult Finnish twins. *Behavior genetics* 2000; **30:** 477-485.

63. Bielicki T, Charzewski J: Body height and upward social mobility. *Annals of human biology* 1983; **10:** 403-408.

64. Cernerud L: Height and social mobility. A study of the height of 10 year olds in relation to socio-economic background and type of formal schooling. *Scandinavian journal of social medicine* 1995; **23:** 28-31.

65. Lasker GW, Mascie-Taylor CG: Effects of social class differences and social mobility on growth in height, weight and body mass index in a British cohort. *Annals of human biology* 1989; **16:** 1-8.

66. Dannenmaier WD, Thumin FJ: Authority Status as a Factor in Perceptual Distortion of Size. *The Journal of social psychology* 1964; **63:** 361-365.

67. Wilson PR: Perceptual distortion of height as a function of ascribed academic status. *The Journal of social psychology* 1968; **74:** 97-102.

68. Hensley WE: Height as a basis for interpersonal attraction. *Adolescence* 1994; **29:** 469-474.

69. Nettle D: Height and reproductive success in a cohort of british men. *Hum Nature* 2002; **13:** 473-491.

70. Pawlowski B, Dunbar RI, Lipowicz A: Tall men have more reproductive success. *Nature* 2000; **403:** 156.

71. Mueller U, Mazur A: Evidence of unconstrained directional selection for male tallness. *Behav Ecol Sociobiol* 2001 **50:** 302–311.

72. Nettle D: Women's height, reproductive success and the evolution of sexual dimorphism in modern humans. *Proceedings* 2002; **269:** 1919-1923.

73. Mascie-Taylor CG: Assortative mating in a contemporary British population. *Annals of human biology* 1987; **14:** 59-68.

74. Silventoinen K, Kaprio J, Lahelma E, Viken RJ, Rose RJ: Assortative mating by body height and BMI: Finnish twins and their spouses. *Am J Human Biol* 2003; **15:** 620-627.

75. Silventoinen K, Sammalisto S, Perola M *et al*: Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 2003; **6:** 399-408.

76. Cavalli-Sforza LL: *African Pygmies*: Academic Press, 1986.

77. Statistics ABo. How Australians Measure Up, 1995.

78. Godoy R, Goodman E, Levins R, Leonard WR: Anthropometric variability in the USA: 1971-2002. *Annals of human biology* 2005; **32:** 469-486.

79. Lasker GW: Migration and physical differentiation: A comparison of immigrant with American-born Chinese. *American Journal of Physical Anthropology* 1946; **4:** 273-300.

80. Smith PK, Bogin B, Varela-Silva MI, Loucky J: Economic and anthropological assessments of the health of children in Maya immigrant families in the US. *Economics and human biology* 2003; **1:** 145-160.

81. Suski PM: The Body Build of American-Born Japanese Children. *Biometrika* 1933; **25:** 323-352.

82. McPherson JD, Marra M, Hillier L *et al*: A physical map of the human genome. *Nature* 2001; **409:** 934-941.

83. Kazazian HH, Jr.: Mobile elements: drivers of genome evolution. *Science* 2004; **303:** 1626-1632.

84. Nachman MW, Crowell SL: Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000; **156:** 297-304.

85. Kong A, Gudbjartsson DF, Sainz J *et al*: A high-resolution recombination map of the human genome. *Nature genetics* 2002; **31:** 241-247.

86. Sherry ST, Ward MH, Kholodov M *et al*: dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 2001; **29:** 308-311.

87. Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. *Nature* 2006; **444:** 444-454.

88. Inoue K, Lupski JR: Molecular mechanisms for genomic disorders. *Annual review of genomics and human genetics* 2002; **3:** 199-242.

89. Sharp AJ, Locke DP, McGrath SD *et al*: Segmental duplications and copy-number variation in the human genome. *American journal of human genetics* 2005; **77:** 78-88.

90. Bailey JN, Almasy L: A brute force dichotomization approach to quantitative trait linkage analysis. *Genetic epidemiology* 1995; **12:** 719-722.

91. Duggirala R, Williams JT, Williams-Blangero S, Blangero J: A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genetic epidemiology* 1997; **14:** 987-992.

92. Galton F: Kinship and correlation. *North American Review* 1890; **150:** 419-431.

93. Galton F: The average contribution of each several ancestor to the total heritage of offspring. *Proceedings of the Royal Society, London* 1897; **LXI:** 401-413.

94. Pearson K: The law of ancestral heredity. *Biometrika* 1903; **2:** 211-236.

95. Fisher RA: The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 1918; **52:** 399-433.

96. Pearson KL, A: On the correlations of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, A* 1901; **195:** 1-47.

97. Falconer DSM, T.F.C: *Introduction to quantitative genetics*, fourth edition edn: Harlow: Longman, 1996.

98. Eaves LJ: Inferring the causes of human variation. *Journal of the Royal Statistical Society, Series A* 1977; **140:** 324-355.

99. Jinks JL, Fulker DW: Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological bulletin* 1970; **73:** 311-349.

100. Amos CI, de Andrade M: Genetic linkage methods for quantitative traits. *Statistical methods in medical research* 2001; **10:** 3-25.

101. Feingold E: Methods for linkage analysis of quantitative trait loci in humans. *Theor Popul Biol* 2001; **60:** 167-180.

102. Lebrec J, Putter H, Houwelingen JC: Score test for detecting linkage to complex traits in selected samples. *Genetic epidemiology* 2004; **27:** 97-108.

103. Putter H, Sandkuijl LA, van Houwelingen JC: Score test for detecting linkage to quantitative traits. *Genetic epidemiology* 2002; **22:** 345-355.

104. Chen WM, Broman KW, Liang KY: Quantitative trait linkage analysis by generalized estimating equations: unification of variance components and Haseman-Elston regression. *Genetic epidemiology* 2004; **26:** 265-272.

105. Daw EW, Heath SC, Wijsman EM: Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer disease pedigrees. *American journal of human genetics* 1999; **64:** 839-851.

106. Heath SC: Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American journal of human genetics* 1997; **61:** 748-760.

107. Lee JK, Thomas DC: Performance of Markov chain-Monte Carlo approaches for mapping genes in oligogenic models with an unknown number of loci. *American journal of human genetics* 2000; **67:** 1232-1250.

108. Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behavior genetics* 1972; **2:** 3-19.

109. Elston RC, Buxbaum S, Jacobs KB, Olson JM: Haseman and Elston revisited. *Genetic epidemiology* 2000; **19:** 1-17.

110. Sham PC, Purcell S, Cherny SS, Abecasis GR: Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *American journal of human genetics* 2002; **71:** 238-253.

111. Wang T, Elston RC: A modified revisited Haseman-Elston method to further improve power. *Human heredity* 2004; **57:** 109-116.

112. Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics* 1998; **62:** 1198-1211.

113. Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. *American journal of human genetics* 1994; **54:** 535-543.

114. Fulker DW, Cardon LR: A sib-pair approach to interval mapping of quantitative trait loci. *American journal of human genetics* 1994; **54:** 1092-1103.

115. Goldgar DE: Multipoint analysis of human quantitative genetic variation. *American journal of human genetics* 1990; **47:** 957-967.

116. Schork NJ: Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *American journal of human genetics* 1993; **53:** 1306-1319.

117. Ekstrom CT, Dalgaard P: Linkage analysis of quantitative trait loci in the presence of heterogeneity. *Human heredity* 2003; **55:** 16-26.

118. Shete S, Zhou X, Amos CI: Genomic imprinting and linkage test for quantitative-trait Loci in extended pedigrees. *American journal of human genetics* 2003; **73:** 933-938.

119. Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J: Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *American journal of human genetics* 1999; **65:** 531-544.

120. Blangero J, Williams JT, Almasy L: Robust LOD scores for variance component-based linkage analysis. *Genetic epidemiology* 2000; **19 Suppl 1:** S8-14.

121. Blangero J, Williams JT, Almasy L: Variance component methods for detecting complex trait loci. *Adv Genet* 2001; **42:** 151-181.

122. Barber MJ, Cordell HJ, MacGregor AJ, Andrew T: Gamma regression improves Haseman-Elston and variance components linkage analysis for sib-pairs. *Genetic epidemiology* 2004; **26:** 97-107.

123. Good PI: *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, 3rd edn: Springer, 2004.

124. Westfall PHY, S.S.: *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*: John Wiley & Sons, 1993.

125. Liu YZ, Xiao P, Guo YF *et al*: Genetic linkage of human height is confirmed to 9q22 and Xq24. *Hum Genet* 2006; **119:** 295-304.

126. Wu X, Cooper RS, Boerwinkle E *et al*: Combined analysis of genomewide scans for adult height: results from the NHLBI Family Blood Pressure Program. *Eur J Hum Genet* 2003; **11:** 271-274.

127. Garn SM, Bailey SM, Cole PE: Similarities between parents and their adopted children. *American journal of physical anthropology* 1976; **45:** 539-543.

128. Langinvainio H, Koskenvuo M, Kaprio J, Sistonen P: Finnish twins reared apart. II: Validation of zygosity, environmental dissimilarity and weight and height. *Acta geneticae medicae et gemellologiae* 1984; **33:** 251-258.

129. Pedersen NL, Friberg L, Floderus-Myrhed B, McClearn GE, Plomin R: Swedish early separated twins: identification and characterization. *Acta geneticae medicae et gemellologiae* 1984; **33:** 243-250.

130. Weedon MN, Lettre G, Freathy RM *et al*: A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature genetics* 2007; **39:** 1245-1250.

131. Battista S, Fidanza V, Fedele M *et al*: The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis. *Cancer research* 1999; **59:** 4793-4797.

132. Ligon AH, Moore SD, Parisi MA *et al*: Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas. *American journal of human genetics* 2005; **76:** 340-348.

133. Zhou X, Benson KF, Ashar HR, Chada K: Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* 1995; **376:** 771-774.

134. Hirschhorn JN, Lindgren CM, Daly MJ *et al*: Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. *American journal of human genetics* 2001; **69:** 106-116.

135. Ellis JA, Scurrah KJ, Duncan AE, Lamantia A, Byrnes GB, Harrap SB: Comprehensive multi-stage linkage analyses identify a locus for adult height on chromosome 3p in a healthy Caucasian population. *Hum Genet* 2006.

136. Wiltshire S, Frayling TM, Hattersley AT *et al*: Evidence for linkage of stature to chromosome 3p26 in a large U.K. Family data set ascertained for type 2 diabetes. *American journal of human genetics* 2002; **70:** 543-546.

137. Liu YZ, Xu FH, Shen H *et al*: Genetic dissection of human stature in a large sample of multiplex pedigrees. *Annals of human genetics* 2004; **68:** 472-488.

138. Willemsen G, Boomsma DI, Beem AL, Vink JM, Slagboom PE, Posthuma D: QTLs for height: results of a full genome scan in Dutch sibling pairs. *Eur J Hum Genet* 2004.

139. Deng HW, Deng H, Liu YJ *et al*: A genomewide linkage scan for quantitative-trait loci for obesity phenotypes. *American journal of human genetics* 2002; **70:** 1138-1151.

140. Geller F, Dempfle A, Gorg T: Genome scan for body mass index and height in the Framingham Heart Study. *BMC Genet* 2003; **4 Suppl 1:** S91.

141. Perola M, Ohman M, Hiekkalinna T *et al*: Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *American journal of human genetics* 2001; **69:** 117-123.

142. Beck SR, Brown WM, Williams AH, Pierce J, Rich SS, Langefeld CD: Age-stratified QTL genome scan analyses for anthropometric measures. *BMC Genet* 2003; **4 Suppl 1:** S31.

143. Mukhopadhyay N, Weeks DE: Linkage analysis of adult height with parent-of-origin effects in the Framingham Heart Study. *BMC Genet* 2003; **4 Suppl 1:** S76.

144. Sale MM, Freedman BI, Hicks PJ *et al*: Loci contributing to adult height and body mass index in African American families ascertained for type 2 diabetes. *Annals of human genetics* 2005; **69:** 517-527.

145. Thompson DB, Ossowski V, Janssen RC, Knowler WC, Bogardus C: Linkage between stature and a region on chromosome 20 and analysis of a candidate gene, bone morphogenetic protein 2. *American journal of medical genetics* 1995; **59:** 495-500.

146. Hasegawa Y, Fujii K, Yamada M *et al*: Identification of novel human GH-1 gene polymorphisms that are associated with growth hormone secretion and height. *The Journal of clinical endocrinology and metabolism* 2000; **85:** 1290-1295.

147. Maheshwari HG, Silverman BL, Dupuis J, Baumann G: Phenotype and genetic analysis of a syndrome caused by an inactivating mutation in the growth hormone-releasing hormone receptor: Dwarfism of Sindh. *The Journal of clinical endocrinology and metabolism* 1998; **83:** 4065-4074.

148. Mullis PE: Genetic control of growth. *Eur J Endocrinol* 2005; **152:** 11-31.

149. Amselem S, Duquesnoy P, Attree O *et al*: Laron dwarfism and mutations of the growth hormone-receptor gene. *The New England journal of medicine* 1989; **321:** 989-995.

150. Kofoed EM, Hwa V, Little B *et al*: Growth hormone insensitivity associated with a STAT5b mutation. *The New England journal of medicine* 2003; **349:** 1139-1147.

151. Takahashi Y, Kaji H, Okimura Y, Goji K, Abe H, Chihara K: Short stature caused by a mutant growth hormone with an antagonistic effect. *Endocrine journal* 1996; **43 Suppl:** S27-32.

152. Woods KA, Camacho-Hubner C, Savage MO, Clark AJ: Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *The New England journal of medicine* 1996; **335:** 1363-1367.

153. Sposito M, Truffarelli F, Sabalich I *et al*: [IGFBP3 in the assessment of growth hormone defects]. *Pediatr Med Chir* 1993; **15:** 67-71.

154. Domene HM, Bengolea SV, Martinez AS *et al*: Deficiency of the circulating insulin-like growth factor system associated with inactivation of the acid-labile subunit gene. *The New England journal of medicine* 2004; **350:** 570-577.

155. Abuzzahab MJ, Schneider A, Goddard A *et al*: IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *The New England journal of medicine* 2003; **349:** 2211-2222.

156. Walenkamp MJ, Wit JM: Genetic disorders in the growth hormone - insulin-like growth factor-I axis. *Horm Res* 2006; **66:** 221-230.

157. Lettre G, Butler JL, Ardlie KG, Hirschhorn JN: Common genetic variation in eight genes of the GH/IGF1 axis does not contribute to adult height variation. *Hum Genet* 2007; **122:** 129-139.

158. Platt JE, Friedhoff AJ, Broman SH, Bond RN, Laska E, Lin SP: Effects of prenatal exposure to neuroleptic drugs on children's growth. *Neuropsychopharmacology* 1988; **1:** 205-212.

159. Arinami T, Iijima Y, Yamakawa-Kobayashi K *et al*: Supportive evidence for contribution of the dopamine D2 receptor gene to heritability of stature: linkage and association studies. *Annals of human genetics* 1999; **63:** 147-151.

160. Smith EP, Boyd J, Frank GR *et al*: Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *The New England journal of medicine* 1994; **331:** 1056-1061.

161. Morishima A, Grumbach MM, Simpson ER, Fisher C, Qin K: Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens. *The Journal of clinical endocrinology and metabolism* 1995; **80:** 3689-3698.

162. Simpson ER: Genetic mutations resulting in loss of aromatase activity in humans and mice. *Journal of the Society for Gynecologic Investigation* 2000; **7:** S18-21.

163. Tiulpakov A, Kalintchenko N, Semitcheva T *et al*: A potential rearrangement between CYP19 and TRPM7 genes on chromosome 15q21.2 as a cause of aromatase excess syndrome. *The Journal of clinical endocrinology and metabolism* 2005; **90:** 4184-4190.

164. Superti-Furga A, Bonafe L, Rimoin DL: Molecular-pathogenetic classification of genetic disorders of the skeleton. *American journal of medical genetics* 2001; **106:** 282-293.

165. Kant SG, Wit JM, Breuning MH: Genetic analysis of short stature. *Horm Res* 2003; **60:** 157-165.

166. Dempfle A, Wudy SA, Saar K *et al*: Evidence for involvement of the vitamin D receptor gene in idiopathic short stature via a genome-wide linkage study and subsequent association studies. *Human molecular genetics* 2006; **15:** 2772-2783.

167. Minamitani K, Takahashi Y, Minagawa M, Yasuda T, Niimi H: Difference in height associated with a translation start site polymorphism in the vitamin D receptor gene. *Pediatric research* 1998; **44:** 628-632.

168. Xiong DH, Xu FH, Liu PY *et al*: Vitamin D receptor gene polymorphisms are linked to and associated with adult height. *Journal of medical genetics* 2005; **42:** 228-234.

169. Deng HW, Li J, Li JL *et al*: Change of bone mass in postmenopausal Caucasian women with and without hormone replacement therapy is associated with vitamin D receptor and estrogen receptor genotypes. *Hum Genet* 1998; **103:** 576-585.

170. Fang Y, van Meurs JB, Rivadeneira F *et al*: Vitamin D receptor gene haplotype is associated with body height and bone size. *The Journal of clinical endocrinology and metabolism* 2007; **92:** 1491-1501.

171. Garnero P, Munoz F, Borel O, Sornay-Rendu E, Delmas PD: Vitamin D receptor gene polymorphisms are associated with the risk of fractures in postmenopausal women, independently of bone mineral density. *The Journal of clinical endocrinology and metabolism* 2005; **90:** 4829-4835.

172. Grundberg E, Brandstrom H, Ribom EL, Ljunggren O, Kindmark A, Mallmin H: A poly adenosine repeat in the human vitamin D receptor gene is associated with bone mineral density in young Swedish women. *Calcified tissue international* 2003; **73:** 455-462.

173. Binder G, Schwarze CP, Ranke MB: Identification of short stature caused by SHOX defects and therapeutic effect of recombinant human growth hormone. *The Journal of clinical endocrinology and metabolism* 2000; **85:** 245-249.

174. Rappold GA, Fukami M, Niesler B *et al*: Deletions of the homeobox gene SHOX (short stature homeobox) are an important cause of growth failure in children with short stature. *The Journal of clinical endocrinology and metabolism* 2002; **87:** 1402-1406.

175. Daw EW, Thompson EA, Wijsman EM: Bias in multipoint linkage analysis arising from map misspecification. *Genetic epidemiology* 2000; **19:** 366-380.

176. Goring HH, Terwilliger JD: Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *American journal of human genetics* 2000; **66:** 1298-1309.

177. Halpern J, Whittemore AS: Multipoint linkage analysis. A cautionary note. *Human heredity* 1999; **49:** 194-196.

178. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American journal of human genetics* 1998; **63:** 861-869.

179. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: GRR: graphical representation of relationship errors. *Bioinformatics (Oxford, England)* 2001; **17:** 742-743.

180. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American journal of human genetics* 1998; **63:** 259-266.

181. Abecasis GR, Cherny SS, Cardon LR: The impact of genotyping error on family-based analysis of quantitative traits. *Eur J Hum Genet* 2001; **9:** 130-134.

182. Douglas JA, Boehnke M, Lange K: A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *American journal of human genetics* 2000; **66:** 1287-1297.

183. Douglas JA, Skol AD, Boehnke M: Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *American journal of human genetics* 2002; **70:** 487-495.

184. Gordon D, Leal SM, Heath SC, Ott J: An analytic solution to single nucleotide polymorphism error-detection rates in nuclear families: implications for study design. *Pacific Symposium on Biocomputing* 2000**:** 663-674.

185. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* 2002; **30:** 97-101.

186. Wigginton JE, Abecasis GR: PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics (Oxford, England)* 2005; **21:** 3445-3447.

187. Barnholtz JS, de Andrade M, Page GP, King TM, Peterson LE, Amos CI: Assessing linkage of monoamine oxidase B in a genome-wide scan using a univariate variance components approach. *Genetic epidemiology* 1999; **17 Suppl 1:** S49-54.

188. Sham PC, Zhao JH, Cherny SS, Hewitt JK: Variance-Components QTL linkage analysis of selected and non-normal samples: conditioning on trait values. *Genetic epidemiology* 2000; **19 Suppl 1:** S22-28.

189. Lander ES, Green P: Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences of the United States of America* 1987; **84:** 2363-2367.

190. Markianos K, Daly MJ, Kruglyak L: Efficient multipoint linkage analysis through reduction of inheritance space. *American journal of human genetics* 2001; **68:** 963-977.

191. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: Allegro, a new computer program for multipoint linkage analysis. *Nature genetics* 2000; **25:** 12-13.

192. Zeegers M, Rijsdijk F, Sham P: Adjusting for covariates in variance components QTL linkage analysis. *Behavior genetics* 2004; **34:** 127-133.

193. Mukhopadhyay N, Finegold DN, Larson MG, Cupples LA, Myers RH, Weeks DE: A genome-wide scan for loci affecting normal adult height in the Framingham Heart Study. *Human heredity* 2003; **55:** 191-201.

194. Hiekkalinna T, Terwilliger JD, Sammalisto S, Peltonen L, Perola M: AUTOGSCAN: powerful tools for automated genome-wide linkage and linkage disequilibrium analysis. *Twin Res Hum Genet* 2005; **8:** 16-21.

195. Rinn JL, Snyder M: Sexual dimorphism in mammalian gene expression. *Trends Genet* 2005; **21:** 298-305.

196. Weiss LA, Pan L, Abney M, Ober C: The sex-specific genetic architecture of quantitative traits in humans. *Nature genetics* 2006; **38:** 218-222.

197. Visscher PM, Hopper JL: Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. *Annals of human genetics* 2001; **65:** 583-601.

198. Song KK, Weeks DE, Sobel E, Feingold E: Efficient simulation of P values for linkage analysis. *Genetic epidemiology* 2004; **26:** 88-96.

199. Wigginton JE, Abecasis GR: An evaluation of the replicate pool method: quick estimation of genome-wide linkage peak p-values. *Genetic epidemiology* 2006; **30:** 320-332.

200. Wiltshire S, Cardon LR, McCarthy MI: Evaluating the results of genomewide linkage scans of complex traits by locus counting. *American journal of human genetics* 2002; **71:** 1175-1182.

201. Wilson EB: Probable inference, the law of succession, and statistical inference. *J Am Statist Assoc* 1927; **11:** 241-247.

202. Silventoinen K: Determinants of variation in adult body height. *Journal of biosocial science* 2003; **35:** 263-285.

203. Rudolf MC, Hochberg Z: Are boys more vulnerable to psychosocial growth retardation? *Developmental medicine and child neurology* 1990; **32:** 1022-1025.

204. Stini WA: Nutritional stress and growth: sex difference in adaptive response. *American journal of physical anthropology* 1969; **31:** 417-426.

205. Stinson S: Sex differences in environmental sensitivity during growth and development: *Physical Anthropology Yearbook 28*. New York: Alan R. Liss, 1985, vol 28, pp 123-147.

206. Silventoinen K, Kaprio J, Lahelma E, Viken RJ, Rose RJ: Sex differences in genetic and environmental factors contributing to body-height. *Twin Res* 2001; **4:** 25-29.

207. Avery CL, Freedman BI, Kraja AT *et al*: Genotype-by-sex interaction in the aetiology of type 2 diabetes mellitus: support for sex-specific quantitative trait loci in Hypertension Genetic Epidemiology Network participants. *Diabetologia* 2006; **49:** 2329-2336.

208. Ober C, Pan L, Phillips N, Parry R, Kurina LM: Sex-specific genetic architecture of asthma-associated quantitative trait loci in a founder population. *Current allergy and asthma reports* 2006; **6:** 241-246.

209. Richards AJ, Yates JR, Williams R *et al*: A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in alpha 1 (XI) collagen. *Human molecular genetics* 1996; **5:** 1339-1343.

210. Griffith AJ, Sprunger LK, Sirko-Osadsa DA, Tiller GE, Meisler MH, Warman ML: Marshall syndrome associated with a splicing defect at the COL11A1 locus. *American journal of human genetics* 1998; **62:** 816-823.

211. Li Y, Lacerda DA, Warman ML *et al*: A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. *Cell* 1995; **80:** 423-430.

212. Benz K, Breit S, Lukoschek M, Mau H, Richter W: Molecular analysis of expansion, differentiation, and growth factor treatment of human chondrocytes identifies differentiation markers and growth-related genes. *Biochem Biophys Res Commun* 2002; **293:** 284-292.

213. Lange K, Sinsheimer JS, Sobel E: Association testing with Mendel. *Genetic epidemiology* 2005; **29:** 36-50.

214. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS: Replication of linkage studies of complex traits: an examination of variation in location estimates. *American journal of human genetics* 1999; **65:** 876-884.

215. Bartels CF, Bukulmez H, Padayatti P *et al*: Mutations in the transmembrane natriuretic peptide receptor NPR-B impair skeletal growth and cause acromesomelic dysplasia, type Maroteaux. *American journal of human genetics* 2004; **75:** 27-34.

216. Olney RC, Bukulmez H, Bartels CF *et al*: Heterozygous mutations in natriuretic peptide receptor-B (NPR2) are associated with short stature. *The Journal of clinical endocrinology and metabolism* 2006; **91:** 1229-1232.

217. Lesnik Oberstein SA, Kriek M, White SJ *et al*: Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase. *American journal of human genetics* 2006; **79:** 562-566.

218. Patel H, Cross H, Proukakis C *et al*: SPG20 is mutated in Troyer syndrome, an hereditary spastic paraplegia. *Nature genetics* 2002; **31:** 347-348.

219. Jaeken J, Wadman SK, Duran M *et al*: Adenylosuccinase deficiency: an inborn error of purine nucleotide synthesis. *European journal of pediatrics* 1988; **148:** 126-131.

220. Varon R, Vissinga C, Platzer M *et al*: Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 1998; **93:** 467-476.

221. Ludecke HJ, Schaper J, Meinecke P *et al*: Genotypic and phenotypic spectrum in tricho-rhino-phalangeal syndrome types I and III. *American journal of human genetics* 2001; **68:** 81-91.

222. Wuyts W, Van Hul W: Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes. *Human mutation* 2000; **15:** 220-227.

223. Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007; **81:** 559-575.

224. Wozney JM, Rosen V, Celeste AJ *et al*: Novel regulators of bone formation: molecular clones and activities. *Science* 1988; **242:** 1528-1534.

225. Salomons GS, van Dooren SJ, Verhoeven NM *et al*: X-linked creatine-transporter gene (SLC6A8) defect: a new creatine-deficiency syndrome. *American journal of human genetics* 2001; **68:** 1497-1500.

226. Marziliano N, Mannarino S, Nespoli L *et al*: Barth syndrome associated with compound hemizygosity and heterozygosity of the TAZ and LDB3 genes. *Am J Med Genet A* 2007; **143:** 907-915.

227. Cabezas DA, Slaugh R, Abidi F *et al*: A new X linked mental retardation (XLMR) syndrome with short stature, small testes, muscle wasting, and tremor localises to Xq24-q25. *Journal of medical genetics* 2000; **37:** 663-668.

228. Hamel BC, Kremer H, Wesby-van Swaay E *et al*: A gene for nonspecific X-linked mental retardation (MRX41) is located in the distal segment of Xq28. *American journal of medical genetics* 1996; **64:** 131-133.

229. Raynaud M, Ronce N, Ayrault AD, Francannet C, Malpuech G, Moraine C: X-linked mental retardation with isolated growth hormone deficiency is mapped to Xq22-Xq27.2 in one family. *American journal of medical genetics* 1998; **76:** 255-261.

230. Vitale E, Specchia C, Devoto M *et al*: Novel X-linked mental retardation syndrome with short stature maps to Xq24. *American journal of medical genetics* 2001; **103:** 1-8.

231. Hughes-Benzie RM, Pilia G, Xuan JY *et al*: Simpson-Golabi-Behmel syndrome: genotype/phenotype analysis of 18 affected males from 7 unrelated families. *American journal of medical genetics* 1996; **66:** 227-234.

232. Gleghorn L, Ramesar R, Beighton P, Wallis G: A mutation in the variable repeat region of the aggrecan gene (AGC1) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis. *American journal of human genetics* 2005; **77:** 484-490.

233. Smits P, Li P, Mandel J *et al*: The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Developmental cell* 2001; **1:** 277-290.

234. Lefebvre V, Li P, de Crombrugghe B: A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *The EMBO journal* 1998; **17:** 5718-5733.

235. Vandenberg P, Khillan JS, Prockop DJ, Helminen H, Kontusaari S, Ala-Kokko L: Expression of a partially deleted gene of human type II procollagen (COL2A1) in transgenic mice produces a chondrodysplasia. *Proceedings of the National Academy of Sciences of the United States of America* 1991; **88:** 7640-7644.

236. Cody JD, Hale DE, Brkanac Z, Kaye CI, Leach RJ: Growth hormone insufficiency associated with haploinsufficiency at 18q23. *American journal of medical genetics* 1997; **71:** 420-425.

237. Bauer FE, Ginsberg L, Venetikou M, MacKay DJ, Burrin JM, Bloom SR: Growth hormone release in man induced by galanin, a new hypothalamic peptide. *Lancet* 1986; **2:** 192-195.

238. Kennedy AM, Inada M, Krane SM *et al*: MMP13 mutation causes spondyloepimetaphyseal dysplasia, Missouri type (SEMD(MO). *The Journal of clinical investigation* 2005; **115:** 2832-2842.

239. Emtage P, Vatta P, Arterburn M *et al*: IGFL: A secreted family with conserved cysteine residues and similarities to the IGF superfamily. *Genomics* 2006; **88:** 513-520.

240. Hsueh WC, Goring HH, Blangero J, Mitchell BD: Replication of linkage to quantitative trait loci: variation in location and magnitude of the lod score. *Genetic epidemiology* 2001; **21 Suppl 1:** S473-478.

241. Curtis D: Genetic dissection of complex traits. *Nature genetics* 1996; **12:** 356-358.

242. Witte JS, Elston RC, Schork NJ: Genetic dissection of complex traits. *Nature genetics* 1996; **12:** 355-356; author reply 357-358.

243. Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature genetics* 1995; **11:** 241-247.

244. Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 2000; **405:** 847-856.

245. Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? *Human molecular genetics* 2002; **11:** 2417-2423.

246. Rogers J, Mahaney MC, Almasy L, Comuzzie AG, Blangero J: Quantitative trait linkage mapping in anthropology. *American journal of physical anthropology* 1999; **Suppl 29:** 127-151.

247. Weiss KM, Terwilliger JD: How many diseases does it take to map a gene with SNPs? *Nature genetics* 2000; **26:** 151-157.

248. Visscher PM, Macgregor S, Benyamin B *et al*: Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American journal of human genetics* 2007; **81:** 1104-1110.

249. Edwards MD, Stuber CW, Wendel JF: Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 1987; **116:** 113-125.

250. Mackay TF: The nature of quantitative genetic variation revisited: lessons from Drosophila bristles. *Bioessays* 1996; **18:** 113-121.

251. Paterson AH, Damon S, Hewitt JD *et al*: Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 1991; **127:** 181-197.

252. Shrimpton AE, Robertson A: The Isolation of Polygenic Factors Controlling Bristle Score in Drosophila Melanogaster. II. Distribution of Third Chromosome Bristle Effects within Chromosome Sections. *Genetics* 1988; **118:** 445-459.

253. Sing CF, Boerwinkle EA: Genetic architecture of inter-individual variability in apolipoprotein, lipoprotein and lipid phenotypes. *Ciba Foundation symposium* 1987; **130:** 99-127.

254. Kearsey MJ, Farquhar AG: QTL analysis in plants; where are we now? *Heredity* 1998; **80 (Pt 2):** 137-142.

255. Hayes B, Goddard ME: The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* 2001; **33:** 209-229.

256. Grandillo S, Tanksley SD: QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species Lycopersicon pimpinellifolium. *Theoretical and Applied Genetics* 1996; **92:** 935-951.

257. Bost B, de Vienne D, Hospital F, Moreau L, Dillmann C: Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics* 2001; **157:** 1773-1787.

258. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *American journal of human genetics* 2001; **69:** 936-950.

259. Evans DM, Cardon LR: Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. *American journal of human genetics* 2004; **75:** 687-692.

260. Kruglyak L: The use of a genetic map of biallelic markers in linkage studies. *Nature genetics* 1997; **17:** 21-24.

261. Sawcer SJ, Maranian M, Singlehurst S *et al*: Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Human molecular genetics* 2004; **13:** 1943-1949.

262. Peltonen L, Palotie A, Lange K: Use of population isolates for mapping complex traits. *Nature reviews* 2000; **1:** 182-190.

263. de la Chapelle A: Disease gene mapping in isolated human populations: the example of Finland. *Journal of medical genetics* 1993; **30:** 857-865.

264. Blangero J: Localization and identification of human quantitative trait loci: king harvest has surely come. *Current opinion in genetics & development* 2004; **14:** 233-240.

265. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* 2006; **38:** 904-909.

266. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155:** 945-959.

267. Goring HH, Curran JE, Johnson MP *et al*: Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature genetics* 2007; **39:** 1208-1216.

268. Morley M, Molony CM, Weber TM *et al*: Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004; **430:** 743-747.

269. Dixon AL, Liang L, Moffatt MF *et al*: A genome-wide association study of global gene expression. *Nature genetics* 2007; **39:** 1202-1207.

270. Stranger BE, Forrest MS, Clark AG *et al*: Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005; **1:** e78.

271. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA: Mapping complex disease loci in whole-genome association studies. *Nature* 2004; **429:** 446-452.

272. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nature reviews* 2005; **6:** 95-108.

273. Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273:** 1516-1517.

274. Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *American journal of human genetics* 2001; **69:** 124-137.

275. Maller J, George S, Purcell S *et al*: Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature genetics* 2006; **38:** 1055-1059.

276. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; **305:** 869-872.

277. Di Rienzo A: Population genetics models of common diseases. *Current opinion in genetics & development* 2006; **16:** 630-636.

278. Reich DE, Lander ES: On the allelic spectrum of human disease. *Trends Genet* 2001; **17:** 502-510.

279. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT: Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005; **437:** 1365-1369.